TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Subset Simulation and Interpolation

Efficient Reliability Estimation under Model-Dynamics for Complex Civil Engineering Structures

FLORIAN BLANDFORT

Dissertation

1. Gutachter: PROF. DR. JÖRN SASS
Technische Universität Kaiserslautern

2. Gutachter: PROF. DR. DANIEL STRAUB
Technische Universität München

Datum der Disputation: 25. Juni 2021

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

D 386

Abstract

The high complexity of civil engineering structures makes it difficult to satisfactorily evaluate their reliability. However, a good risk assessment of such structures is incredibly important to avert dangers and possible disasters for public life. For this purpose, we need algorithms that reliably deliver estimates for their failure probabilities with high efficiency and whose results enable a better understanding of their reliability. This is a major challenge, especially when dynamics, for example due to uncertainties or time-dependent states, must be included in the model.

The contributions are centered around Subset Simulation, a very popular adaptive Monte Carlo method for reliability analysis in the engineering sciences. It particularly well estimates small failure probabilities in high dimensions and is therefore tailored to the demands of many complex problems. We modify Subset Simulation and couple it with interpolation methods in order to keep its remarkable properties and receive all conditional failure probabilities with respect to one variable of the structural reliability model. This covers many sorts of model dynamics with several model constellations, such as time-dependent modeling, sensitivity and uncertainty, in an efficient way, requiring similar computational demands as a static reliability analysis for one model constellation by Subset Simulation. The algorithm offers many new opportunities for reliability evaluation and can even be used to verify results of Subset Simulation by artificially manipulating the geometry of the underlying limit state in numerous ways, allowing to provide correct results where Subset Simulation systematically fails. To improve understanding and further account for model uncertainties, we present a new visualization technique that matches the extensive information on reliability we get as a result from the novel algorithm.

In addition to these extensions, we are also dedicated to the fundamental analysis of Subset Simulation, partially bridging the gap between theory and results by simulation where inconsistencies exist. Based on these findings, we also extend practical recommendations on selection of the intermediate probability with respect to the implementation of the algorithm and derive a formula for correction of the bias. For a better understanding, we also provide another stochastic interpretation of the algorithm and offer alternative implementations which stick to the theoretical assumptions, typically made in analysis.

Acknowledgments

First and foremost, I would like to express my gratitude to Prof. Dr. Jörn Sass for his constant support and supervision. He was always available for discussion, helped me to improve the quality of my research and gave me valuable feedback on work and progress throughout the whole past three years. I would also like to thank Dr. Stefanie Schwaar, with whom I had several in-depth discussions, making many aspects clearer and thereby significantly contributing to the quality of this work. In particular, I am very grateful for the opportunity to take this journey into the world of reliability estimation of civil engineering structures with support and supervision by both of them.

Also, I am very grateful for the support granted within the RTG 1932 'Stochastic Models for Innovations in the Engineering Science' which is funded by the Deutsche Forschungsgemeinschaft (DFG). In particular, being part of the RTG 1932 allowed me to work on an interdisciplinary topic and getting to know several nice people and interesting topics from other disciplines, which I, apart from the financial support, appreciated very much.

With reference to this, as a business mathematician, I am also very grateful for the members of the group 'Concrete Structures and Structural Design' of the Department of Civil Engineering, which patiently helped me to understand the language and fundamentals of engineering, especially regarding concrete structures. In particular, I thank Rabea Sefrin and Kasem Maryamh which helped me a lot at judging mathematical ideas from an engineering perspective.

Moreover, I would like to thank the members of the Financial Mathematics Group for providing a fruitful working environment and excellent discussions, particularly in the coffee kitchen but also anywhere else.

Last but not least, I want to thank my wife and family, who always give me permanent support and managed to keep me attached to the real world during finishing this dissertation.

Contents

Chapter 1

Introduction

The complexity of numerical models for evaluation of engineering structures is constantly increasing. Although computational power is also rapidly increasing, typically a single analysis can take hours or days (Stocki et al. (2017), Yi et al. (2020)), even on multi-core machines (Bourinet (2018)). Development of highly efficient tools for reliability evaluation is thus crucial for an adequate assessment of engineering structures and still a major challenge. These tools require, beyond a high efficiency, a strong theoretical basis and also verification of their functionality in practice. First, since the complexity of analysis of such structures can make it impossible to decide whether results for real applications are reasonable or not. Second, the consequences of bad assessment can be severe and even cost human lives. This extremely unfavorable combination requires the utmost caution which has led many researchers to contribute to this field and to continuously improve methods for reliability evaluation.

One can either use deterministic methods such as the First Order Reliability Method (Hasofer and Lind (1974)), typically referred to as FORM, and Second Order Reliability Method (SORM) (e.g. Fiessler et al. (1979), Breitung (1984), Tvedt (1990)) or approach the reliability evaluation by Monte Carlo methods. High dimensional integration and highly non-linear or only implicitly given limit state functions (Schuëller and Pradlwarter (2007), Au and Wang (2014), Su et al. (2017), Chehade and Younes (2020)) make reliability analysis of complex structures particularly challenging. The general applicability of Monte Carlo methods, especially for highly non-linear limit state functions, where FORM and/or SORM lead to wrong results (see e.g. Zhao and Ono (1999), Grandhi and Wang (1999), Bourinet et al. (2011)),

high-dimensional problems (Schuëller et al. (2004)) or shortcomings associated with multiple design points which are hard to overcome (Der Kiureghian and Dakessian (1998),Zhao and Ono (2001)), and the technological advances, have made the Monte Carlo approach an important tool in structural reliability analysis. Due to its high computational costs when exploring small failure probabilities, several methods to improve its efficiency have been introduced (see e.g. Engelund and Rackwitz (1993), Koutsourelakis et al. (2004)) and enhanced in various directions. Popular methods are, among others, Subset Simulation (SuS) (Au and Beck (2001b), Ching et al. (2005), Papaioannou et al. (2015), Au and Patelli (2016)), Line Sampling (Koutsourelakis et al. (2004), De Angelis et al. (2015)) or a Monte Carlo based extrapolation technique developed in Bucher (2009) (also compare Naess et al. (2009), Qin et al. (2012)). These methods were originally developed to evaluate the failure probability with respect to structures with given stochastic properties and expected demands. Assuming a fixed model, the algorithms rely on importance sampling schemes where failure probabilities are calculated by exploring specific regions of the variable space, often accurately delivering results for the given probabilistic model, but providing little information about deviations from the model, especially when they are not very small. This is necessary as only small parts of the sample space are relevant for failure and exploring the whole variable space is often infeasible. So, these methods are static by nature as they trade efficiency by locally restricted information for robustness and global information.

Dynamic Models: A Substantial Challenge As opposed to the static nature of Monte Carlo methods relying on importance sampling, the demand for efficient analysis of dynamic models, where several model constellations are relevant, is high for many reasons. First, stochastic properties of engineering structures change over time (see e.g. Stewart and Rosowsky (1998), Czarnecki and Nowak (2008), Li et al. (2015b), Schneider et al. (2017)), requiring an analysis which accounts for the evolution of stochastic properties with respect to time. Situations such that the effects of time-dependent model dynamics result in a non-negligible change of the failure probabilities, frequently appear if structures are deteriorating or if demands, such as traffic volume, change or if inspections yield new information about the structure. Second, uncertainties or sensitivities with respect to the stochastic model (e.g. Song et al. (2009), Guo and Du (2009),Wei et al. (2013), Yun et al.

(2018a)) play a major role in adequate decision making. When we have only little knowledge about the model, it is clear that we have to quantify these uncertainties and check the corresponding consequences. Particularly slightly wrong model assumptions must not propagate and dramatically increase failure probabilities. Also, we should identify the most important impacts on reliability and account for uncertainty propagation and sensitivity of the failure probability with respect to the variables of the model (compare e.g. Wei et al. (2012), Yun et al. (2018b)). Third, in reliability based design optimization (see e.g. Au (2005), Beck and de Santana Gomes (2012), Melchers and Beck (2018), Bourinet (2018)) we need to consider several constellations of the stochastic properties of the structure or sensitivities thereof. Facing a constrained optimization problem also considering the design costs, only allows to draw meaningful conclusions if the corresponding reliabilities can be estimated accurate enough at a feasible computational demand. Observe that the given distinction of dynamic models is not strict and methodologies used in one case can as well serve in another case.

The calculation of such models for complex structures in a feasible time is often difficult. The naive approach of repetitive evaluations for relevant constellations of the model will often result in infeasible computational demands. Therefore, developing methods to particularly address the efficient computation of dynamic models is given increasingly much relevance for quite some time now. However, regarding the evaluation of small failure probabilities, it is a challenging task to keep exploring important regions of the sample space while keeping the necessary number of limit state evaluations low and trying to collect information about different model constellations such as by uncertainties or time-dependence. The reason is that we are now confronted with several important regions of the state space, each belonging to a specific model constellation. These can be close to each other in the state space, but they may also differ substantially. Additionally, it may also be difficult to identify interesting model constellations and those which have to be explored by priority, as relevance also depends on the failure probability which is not available in advance. Current approaches use repetitive calculations for a selected number of constellations (e.g. Schneider et al. (2017)), surrogate models (Bourinet (2018)), determine local sensitivities with respect to distribution parameters (Song et al. (2009)), add uncertainty to get an augmented reliability problem (Au (2005)) to receive more information but with lower resolution, compute boundaries for the failure probability (De Angelis et al. (2015), Alvarez et al. (2018)) or consider interval probabilities (e.g.

De Angelis et al. (2015)). In time-dependent reliability, there are a number of specialized methods to compute failure probabilities. These are, for example, first passage based methods (Andrieu-Renaud et al. (2004), Shi et al. (2017)), extreme value based methods (Li et al. (2007), Du (2014)) or surrogate models (Zhang et al. (2017), Bourinet (2018)) which allow to attain low computational costs but trade accuracy for this efficiency. Surrogate models are promising, but, as they are based on approximations, might also lack accuracy sometimes. Current models may in particular reach their limits in high-dimensional or certain complex problems (Bourinet (2018)).

These approaches work very well in many settings. However, according to the difficulty at exploring several important regions of the sample space, they often require high computational costs or achieve low accuracies for estimated failure probabilities if many constellations are considered and in particular if these differ substantially, for example if distribution types of stochastic variables are changed. This is often inevitable, since crucial modifications will, at some point, make it impossible to derive information from several such models at the same time, so some restrictions are certainly required. Most existing approaches are either focusing on local properties of model dynamics or rely on approximations that may fail in case of complex structures. Even if we get a solution, we often remain uncertain about the correct derivation of the failure probability, as we often rely on Black Box models for computation. Thus, besides the consideration of efficiency in dynamic models, we additionally need to examine the statistical properties of the utilized algorithms and try to include information on unknowns or uncertainties in the model, their effects on the resulting failure probabilities and an adequate presentation thereof.

The high Potential of Subset Simulation In our contribution, we face these three challenges by extending the statistical analysis of Subset Simulation, introducing a novel algorithm for efficient reliability evaluation of dynamic models and presenting a new visualization technique that matches very well to the extensive information on reliability we get as a result from our algorithm. The developments are largely based on the algorithm Subset Simulation or at least on the principles thereof. As a consequence, our proposed new algorithm rather does not compete with approaches for dynamic models of structures for which it is also not beneficial to use Subset Simulation in general. Comparative studies on efficiency were carried out in

Schuëller et al. (2004), Schuëller and Pradlwarter (2007), Seplveda and Faber (2019). We mainly focus on improving the analysis of such structures where Subset Simulation performs good and in addition on those that may not yet be well analyzed in some dynamic settings by algorithms based on it.

The reason for a focus on Subset Simulation and its extension is, asides from its high efficiency, is also as follows. Its incredible popularity across many disciplines suggests that the algorithm defines an important tool for reliability evaluation with remarkable properties. Originally developed for estimating reliability of structures exposed to uncertain earthquake ground motions (Au and Beck (2001b)), it has also been found to suit problems in research fields such as aerospace (Pellissetti et al. (2006)), nuclear (Zio and Pedroni (2012)), and geotechnical engineering (Phoon (2008), Au and Wang (2014)), failure rate of nanoscale circuits (Sun and Li (2014)), estimation of impact probabilities of asteroids (Romano et al. (2020)) and incident prediction in flight safety (Wang et al. (2014)). Recently, on top of these important fields of study, it was also shown that Subset Simulation is an interesting candidate for use in dealing with current challenges: analysis of complex networks such as critical infrastructures (Zuev et al. (2015), Zuev and Beer (2018)) and considerations with respect to climate change (Yang and Wang (2014), Elsheikh et al. (2014), Song and Wang (2016), Xiao et al. (2019)). This is not surprising, because a special characteristic of Subset Simulation is that it suits high-dimensional problems particularly well, so that it is natural to consider it as a good option for analysis of such large and complex systems. One reason for the widespread use of the method is most likely its simplicity, making it easy to understand and implement. Many problems of the future will certainly be interdisciplinary, so the basic knowledge of the algorithm is also advantageous in discussions, where different perspectives and terminology might already be challenging. Another advantage is the information it provides beyond a failure probability estimate. By exploring the failure domain of the variable space with Monte Carlo, it provides states that most likely result in failure of the structure so that critical properties and importance of individual stochastic variables of the structure can be identified. This way, it might provide insights into the mechanics behind failure and thereby enable engineering experts to judge the result afterwards.

On the other hand, apart from its great advantages, there are also some challenges we have to deal with. A very critical one, which has to be considered is given in Breitung (2019). Indeed, it is demonstrated that for specific

structures, Subset Simulation can systematically underestimate the failure probability. This issue has to be given attention, particularly for complex structures where an interpretation of the result is often difficult.

Main Contribution The main contribution of this work is the introduction and study of a novel algorithm 'Subset Simulation Interpolation' (SuSI) in Chapter 5 and Chapter 6. It couples Subset Simulation with interpolation methods and can efficiently evaluate all relevant conditional failure probabilities with respect to one variable of the probabilistic model. Chapter 5 yields the introduction, implementation and statistical analysis of the algorithm and Chapter 6 provides tests by simulation as well as application opportunities. Both approaches for analysis, theoretical and practical, suggest similar computational demands as for an ordinary Subset Simulation evaluation of a static model. In contrast to ordinary Subset Simulation, the algorithm collects direct information about failure probabilities of the structure in every subset to get more information despite similar effort. Thereby, we also receive information about reliability of substantially differing constellations of the probability model.

As briefly discussed in the introduction, limitations regarding the considered dynamic model appear to be inevitable. Therefore, we need to specify a dynamic model, the algorithm will focus on. Our limitation is the reflection of dynamics by one model variable which needs to have a monotone effect on the limit state function and to be independent of all other stochastic model variables. Detailed information on this dynamic model is given in Chapter 2. Although this might seem quite restrictive at first, it yields several non-obvious opportunities regarding dynamic models. Consider for example an artificial time variable that has an effect on many stochastic properties of the structure. Furthermore, it can be capable of solving problems where Subset Simulation yields wrong results. Therefore, we could use it for validation of Subset Simulation results, identifying cases where it systematically fails. This is particularly interesting because both algorithms perform well on the same type of problems so that it could generally be used, not only for derivation of stand-alone results, but also as a backup test and to generally identify structures where Subset Simulation fails. This is a big step towards a solution to the main drawback of ordinary Subset Simulation. These examples for application, and additional ones, are given in Chapter 6 as well as Chapter 8. Also note that additional information on uncertainty or sen-

sitivity can be explored in our new algorithm by applying similar methods as those developed for ordinary Subset Simulation. The same applies to methods for increased efficiency. There is not much direct competition, since the algorithm differs greatly in the information provided and its structure. However, it retains the basic principle and fundamental structure of Subset Simulation, so that the same or similar methods for increased efficiency and information will often apply for it too. So our approach can also be combined with others and provide additional information or efficiency on top of them.

Apart from Subset Simulation Interpolation, we also dedicate ourselves to other topics centered around Subset Simulation, although the main purpose remains extending knowledge related to the new algorithm. First, we shortly review ordinary Subset Simulation, in particular its fundamental analysis and extensions for dynamic models, in Chapter 3. An extension of its stochastic interpretation follows in Chapter 4, dealing with the discrepancies between theory and practice in simulations. This allows for an easier modification of Subset Simulation in meaningful ways, thus helps building and understanding the novel algorithm. Nevertheless, this chapter also provides important stand-alone results for ordinary Subset Simulation. In Chapter 7, we discuss efficient calculation of a specific time-dependent model which can be considered as a particularly efficient special case of our main contribution. Lastly, we present a new visualization technique that perfectly utilizes the extensive information on reliability received in this special case, and also by our novel algorithm, in Chapter 8. A detailed list of all contributions is given in Section 2.3 and the corresponding scope in view of applicability is stated in Section 2.4.

Chapter 2

Scope and Methodology

In this chapter, we introduce the considered dynamic reliability model, show how we measure efficiency of algorithms and provide a list of our contributions to reliability analysis of complex structures.

2.1 Reliability Estimation

Given a structure of any kind, we want to know the probability that it is capable to fulfill the demands for which it was designed for. We would like to remain as general as possible and only if necessary resort to certain special cases, such as with respect to model dynamics where a specification is indispensable. First, we state the general reliability model.

2.1.1 General Reliability Model

Let $\mathbf{X} := (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_d)$ be a continuous d-dimensional real-valued and non-degenerate random variable, where the $\mathbf{X}_i$, $i = 1, ..., d$, are called basic variables (or stochastic properties) of the structure. Their probability density functions (pdfs) are $f_1, f_2, ..., f_d$ while $f_{\mathbf{X}}$ denotes the joint pdf. The basic variables represent properties of the structure as e.g. material properties, geometric features or loads which the structure is exposed to. In this setup, we would like to know the probability that a function $g : D \to \mathbb{R}$ with $D = (D_1, ..., D_d) \subseteq \mathbb{R}^d$, called limit state function, takes a value smaller than

some threshold $b^* \in \mathbb{R}$ given the stochastic properties of the structure:

$$E\left[\mathbf{1}_{\{g(\mathbf{X})<b^*\}}\right] = P(g(\mathbf{X}) < b^*) = \int_D f_{\mathbf{X}}(\mathbf{x})\mathbf{1}_{\{g(\mathbf{x})<b^*\}}d\mathbf{x} \tag{2.1}$$

The case $g(\mathbf{x}) < b^*, \mathbf{x} \in D$ corresponds to failure of the structure. We thus also write $p_f = P(g(\mathbf{X}) < b^*)$ and call it the failure probability. Correspondingly, we call the probability $L = 1 - p_f$ of the complement $\{g(\mathbf{X}) \geq b^*\}$ the reliability of the structure. In the following, we will refer to such a model with fixed constellation (i.e. one evaluation of (2.1)) by the name 'static model'. Many more formulations of reliability models are stated in e.g. Melchers and Beck (2018).

Correlations. When correlations are present, it is convenient to derive a solution by transforming the random variables $\mathbf{X}$ into independent standard normal ones. Typical transformations are Rosenblatt (Hohenbichler and Rackwitz (1981)) or Nataf transformation (Der Kiureghian and Liu (1986)). See Li et al. (2008) or Xiao (2014) for more information on how to efficiently implement a Nataf transformation. Also see e.g. Bourinet (2018) for detailed information on how to apply these transformations in reliability evaluation. As this is the general procedure to deal with correlated random variables, we can equivalently solve the corresponding version of Equation 2.1 in Gaussian space. This transformation will, with only a few exceptions, not be addressed in this work. However, it is important to keep in mind that it can be utilized so that we are not restricted to models with independent variables. Our contribution as well suits cases where correlations are present. Violations of this statement, requiring independence for specific variables, will explicitly be addressed.

2.1.2 Complexity and Efficiency

As our analysis aims at comparing the efficiency of different algorithms, we need to define our indicator for efficiency. Our focus lies on the analysis of complex engineering structures. Thus (2.1) is generally hard to solve and evaluation of it takes a significant amount of time. First, we have to specify the properties of so-called complex structures in this dissertation.

Assumption 2.1.1 (Complex Structures)**.** *In general, we make the following assumptions for* (2.1) *to represent the given complexity (compare Zuev et al. (2015)):*

(S1) There is no explicit formula for $\mathbf{1}_{\{g(\mathbf{x})<b^*\}}$.

(S2) Evaluation of $g(\mathbf{x})$ *for a given realization* $\mathbf{x}$ *of* $\mathbf{X}$ *is considered as the significant computational demand.*

(S3) The dimension d *of* $\mathbf{X}$ *is large (e.g.* $d \geq 100$*).*

(S4) The failure probability p_f *is small. Typically* $p_f \approx 10^{-j}, j = 3, ..., 9$.

These assumptions need not all to be fulfilled for every structure, but we want to guarantee that structures with all these characteristics can be considered in our contributions. This thesis focuses on general concepts and methods for dealing with complex structures. Accordingly, no finite element methods or external programs, which again would be common in analysis of complex structures, are used in our presented examples. Instead, examples with explicit limit state functions are given. This allows for a better traceability of the results and especially to recognize limits of the presented methods, because then more information such as the geometry of the limit state function is available. However, regarding the assumptions, in particular the high dimensionality of $\mathbf{X}$ makes it impossible to solve (2.1) analytically, demanding for specific algorithms which can give approximate results in a feasible computation time. Now every algorithm needs a specific number of total function evaluations E_T of this expensive integral to achieve a specific accuracy. Next, we state how the efficiency of such algorithms will be measured.

Definition 2.1.2 (Measuring Efficiency)**.** *We measure the efficiency of an algorithm by the value of its Mean Squared Error (*MSE(E_T)*) with respect to its total number of limit state evaluations* E_T*. As a representative, we consider the relative root mean squared error (*rRMSE*) with respect to the real failure probability* p_f *for ease of interpretation. Thus, if an algorithm delivers a lower* MSE *(resp.* rRMSE*) than another algorithm for the same number of total function evaluations* E_T *in an application, it is considered more efficient for the corresponding analysis. When pointing out specific properties of the distribution of results by simulation, we will also provide other benchmarks, such as mean absolute error or empirical quantiles, to get a more holistic impression. In some special cases, such an extension of information can be extremely important. Otherwise, it may remain unclear, for example, whether we underestimate or overestimate the probability of failure.*

Counting the number of limit state evaluations, neglecting other computational demands such as generating random variables or performing low dimensional integration, is due to the fact that typically evaluation of g is very expensive and yields the main computational burden (cf. Zuev et al. (2015)) as stated by Assumption (S2). Exceptions are sometimes briefly discussed in our analysis. If no relevant bias exists, the rRMSE naturally reduces to the coefficient of variation (CV). Regarding the measure for efficiency, another approach would be to look at intervals that contain specific values with a given probability. This has the advantage of allowing us to evaluate probabilities of benchmark violations, but requires to assume a distribution that might be unknown. Anyhow, empirical estimations of quantiles in experiments are feasible. Typically we need many assumptions to derive the properties of advanced stochastic algorithms, so that one should combine results from statistical analysis and simulation.

Remark 2.1.3. *It is important to take into account both, theoretical and simulation based results in analysis of advanced algorithms for complex structures to avoid drawing only example based conclusions for specific situations or stand-alone theoretical ones that do possibly not suit practice.*

This is particularly important, since in applications we will typically face Black-Box-like limit state evaluations (compare (S1)), e.g. by a tremendous complexity of underlying numerical computations, so that it is often practically impossible to decide whether results in reliability analysis are reasonable or not. Ensuring the accuracy of the applied methods is therefore extremely important.

2.2 Focus: Dynamic Models

We particularly examine the special case of dynamic models, where the model constellation is not fixed and changes in the underlying stochastic model occur. However, instead of allowing arbitrary dynamics in the model, we choose a dynamic model where all dynamics can be attributed to one stochastic variable of the model. This variable is referred to as 'dynamic variable' $\mathbf{X}_k$. At first, this might seem very restrictive, but in practice this yields the potential to cover many interesting realistic settings while being computational feasible. Situations that can be covered are e.g. knowing a dominant variable of the model, having most uncertainty in one variable or representing

time effects by some time-dependent stochastic process which is part of the limit state function. The opportunities are widespread, also often allowing to approximate even more sophisticated settings by e.g. artificially introduced stochastic variables. Some useful cases will be demonstrated in Chapter 6 and in Chapter 8. Alternatively, also the analysis of robustness or sensitivity of p_f with respect to the choice of the pdf f_k of $\mathbf{X}_k$ or its parameters is captured in the dynamic model this way.

In all these cases, the model (2.1) then needs to be evaluated for several different pdfs f_k of the stochastic variable $\mathbf{X}_k$. We can rewrite the structural reliability equation (2.1) with conditional failure probabilities with respect to the dynamic variable $\mathbf{X}_k$, resulting in

$$p_f = \int_{D_k} P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y) f_k(y) dy \ . \tag{2.2}$$

Since we are interested in dynamic models here, i.e. we want to compute the reliability for several different models of $\mathbf{X}_k$, the assumption of independence of $\mathbf{X}_k$ is made.

Assumption 2.2.1 (Independence)**.** *The basic variable $\mathbf{X}_k$ is independent of all other model variables $\mathbf{X}_i, i \in \{1, ..., d\}$ with $i \neq k$.*

This assumption is necessary to allow computing Equation 2.2 by a simple one dimensional integration for several pdfs f_k, once $P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y)$ is calculated for all relevant $y \in D_k$. To demonstrate the necessity of the assumption of independence of the dynamic variable $\mathbf{X}_k$, an example of a stochastic model with present correlations is given in Appendix A. Otherwise, for different joint distributions the conditional probability $P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y)$ could have different values (compare Appendix A). Furthermore, we will also add another assumption which is not required for the same purpose but instead is utilized in our algorithm 'Subset Simulation Interpolation', which we introduce in this thesis (Chapter 5).

Assumption 2.2.2 (Monotonicity)**.** *The limit state function g is monotone with respect to $\mathbf{X}_k$. Without loss of generality, we further assume $g(\mathbf{x}_1, ..., \mathbf{x}_d)$ is increasing in component $\mathbf{x}_k$.*

The algorithm will successively compute the conditional failure probabilities similar to the procedure given in Subset Simulation for varied thresholds,

so that the monotonicity condition allows to explore the domain of $\mathbf{X}_k$, starting at low deterministic $\mathbf{x}_k$ and successively heightening its value.

Those assumptions are necessary to allow efficient reliability evaluation of dynamic models with the novel algorithm. However, contributions given in Chapter 4 and Chapter 7 deal with the standard static setting also. There are many cases of structural reliability analysis where both assumptions hold. Temporary loads, for example, often fulfill both assumptions naturally, as those are caused by external effects and are often assumed to not interact with any other stochastic property of the structure (neglecting fatigue effects). Alternatively, it can also be beneficial to introduce a new artificial variable (e.g. a time variable), capturing the dynamic effects and fulfilling both assumptions. Indeed, the opportunity to identify an artificial variable as the dynamic variable offers many interesting applications, as demonstrated in Section 6.2.

2.3 Contributions

Our main focus is on dynamic models. However, we do not limit ourselves to reliability analysis of such dynamic models, but instead also examine the fundamental basis of the newly introduced algorithm, ordinary Subset Simulation. This way, we contribute to the understanding of this popular original algorithm, before adding additional complexity (to suit dynamic models) in the presented algorithm. We even benefit directly from our results as for example a more flexible choice of parameter settings in ordinary Subset Simulation is shown to not reduce efficiency, providing more opportunities for modification. Furthermore, a parameter state model for time-dependent reliability evaluation is presented. Some of the contributions have previously been presented and published. The corresponding publications, if existent, are referenced in the following. In detail, we make the following contributions:

- Chapter 3: We give a short overview on fundamental improvements of Subset Simulation, modifications that are capable to deal with similar dynamic models as the one proposed and sum up important challenges ordinary Subset Simulation is confronted with.

- Chapter 4: A novel analysis and thereby derived enhancements of ordinary Subset Simulation are presented. Our aim is to bridge the gap be-

tween theoretic study and practical findings in simulations. This yields an extension of the fundamental basis of Subset Simulation to explore new opportunities for future modifications, such as in our proposed algorithm, and provides information on fundamental understanding of the algorithm:

- We extend the analysis of Subset Simulation by giving a novel stochastic interpretation, thereby further improving understanding of the algorithm.
- It is shown how, under weak conditions, the conditional probabilities in Subset Simulation, which we refer to as intermediate (subset) probabilities, typically desired to be equal to some fixed probability $p_0 \in [0.1, 0.3]$, are theoretically independent. Different results in simulations are discussed, explained and erased by a slight change in the implementation setting that results in a similar computational demand as not re-using seeds in the algorithm. Also, evidence on independence of intermediate subset probabilities is stronger for higher choices of p_0 in the original implementation.
- We show how the definition of the thresholds and seed selection for Markov Chain Monte Carlo (MCMC) in Subset Simulation can be brought in line with theoretical analysis, avoiding uncontrollable effects. This also exposes that there might be a potential to have a negative bias in current implementations of Subset Simulation.
- The chain length in MCMC was found to have a significant effect on the efficiency of Subset Simulation. If the chain length is adequately adapted, we are not restricted to parameters such as $p_0 \in [0.1, 0.3]$, but may instead also use higher p_0, such as $p_0 = 0.5$ or $p_0 = 0.8$ with the same (for very low failure probabilities possibly even slightly higher) efficiency.
- We develop an explicit formula for bias correction of Subset Simulation results which does not require asymptotic arguments. As a consequence, also smaller sample numbers become an interesting field of study, when they have an undesirable bias otherwise.
- In simulations, we provide evidence for the results by stochastic analysis.

- Chapter 5: A new algorithm 'Subset Simulation Interpolation' is presented and analyzed theoretically. This is the first part of the main contribution of the thesis.[1]

  - We propose a novel algorithm, based on Subset Simulation, designed for efficient reliability evaluation under model dynamics. It efficiently computes conditional failure probabilities for arbitrary probability distributions of one stochastic variable of the model.
  - The algorithm combines Subset Simulation with interpolation methods. For interpolation, we use cubic splines as well as approximation by staircase functions.
  - We prove that, under certain conditions, the computational demand of the new algorithm is similar to that of a static reliability evaluation by Subset Simulation. The constructive proof provides a good basic understanding on the mechanics in evaluation by the new algorithm.

- Chapter 6: This chapter extends Chapter 5 by practically studying the new algorithm and showing application opportunities.

  - We demonstrate similarity of the computational demand of the algorithm with respect to ordinary Subset Simulation in simulations, considering different failure probabilities, geometries of the limit state function, dimensions and parameter settings.
  - The algorithm is, as Subset Simulation, particularly good for extensive analysis of complex high-dimensional structures with small failure probabilities. The form of our dynamic model allows for new interesting fields of study because asking and answering more sophisticated or new questions becomes computational feasible or less demanding. Some examples are given for illustration, showing for example an efficient evaluation of time-dependent reliability or an extensive sensitivity analysis with respect to one stochastic variable.

[1] The algorithm was first presented at ESREL2019 (Blandfort et al. (2019a)). Since then it has been further developed, applications were extended and a theoretical foundation was given. The results are presented in this dissertation.

  - Furthermore, we discover that modifications of the limit state function geometry, present in the new algorithm, could be extremely beneficial since many different ways to compute the same failure probability become available, while keeping most useful properties of Subset Simulation. In particular, this does not require specific model assumptions, is supposed to be general applicable and can also deliver correct results in cases where Subset Simulation (SuS) critically underestimates the failure probability.

- Chapter 7: A way to efficiently compute reliability in time-dependent capacity-demand models by conversion of the original model to a parameter state model is presented.[2]

  - Similar to our new algorithm, this approach is based on the representation by conditional failure probabilities with respect to one variable.
  - Using mathematical properties of the model, all time-dependent information is mapped to one variable, allowing to consider many effects by analyzing changes in this one variable.
  - Additionally, an explicit formula for calculation of time-dependent reliability in this model, under consideration of survival of the structure, is given.

- Chapter 8: We show how to illustrate results of dynamic models in a comprehensive and informative way, using a visualization tailored to the information given in our approaches.[3]

  - It is shown how to visualize the results of Chapter 6 and Chapter 7 by heat maps, allowing to improve understanding of effects on reliability by model uncertainties and dynamics.
  - To demonstrate the visualization technique, several examples are given, under dynamic models such as required for application of SuSI and also in the special case where the parameter state model is suitable.

[2]The contents of this chapter are a slightly altered version of Blandfort et al. (2021). Many parts coincide, but we do not introduce the presented visualization technique here and add references to SuSI.

[3]The content of this chapter is partially an extract of Blandfort et al. (2021), containing but also extending the visualization part of it and applying it on SuSI additionally.

  - We consider examples of a concrete slab, a network problem, a lossy transmission line and a two degree of freedom damped oscillator.

- An implementation providing the novel algorithm 'Subset Simulation Interpolation' and the option for extended parameter settings in Subset Simulation, as utilized in Chapter 4 as well as Chapter 7, is publicly available at GitHub: https://github.com/FBlandfort/Subset-Simulation-Interpolation.

2.4 Scope in View of Applicability

All of our considered computational approaches are based on SuS. Thus, they deal with the reliability analysis of complex engineering structures. In particular, SuS efficiently evaluates small failure probabilities in high dimensions, often present in the stochastic models for analysis of complex structures. The scope of this thesis is as follows.

- Ordinary Subset Simulation (Chapter 3, Chapter 4):
  - SuS can (theoretically) be applied on any static reliability model.
  - The given analysis in Chapter 4 is mainly based on modern implementations as described at the beginning of Chapter 4.
  - Contributions in Chapter 4 are specialized on either small and moderate or high sample numbers, but generally apply to ordinary Subset Simulation.

- Subset Simulation Interpolation (Chapter 5, Chapter 6)
  - If used for altering the limit state function only, SuSI can be applied on any static reliability model, just like SuS. Examples are given in Section 6.2.3, where the alteration allows to receive correct results by SuSI where SuS systematically and severely underestimates the true failure probability.
  - Considering dynamic models, SuSI can be applied on any dynamic model such as described in Section 2.2. This dynamic model does not necessarily need to be given in the stochastic structural model directly, but can also be artificially changed into the desired form.

- Parameter State Model (Chapter 7):
  - The parameter state model, as well as SuSI, should be applied on dynamic models, but is much more restrictive in the model compared to SuSI. In the static case, it already requires a capacity-demand ($R - S$) reliability model with independent R and S and demands represented by a single stochastic or deterministic variable. If time-dependent effects are included into the model, this single stochastic variable S must additionally belong to the location-scale family (e.g. Gaussian, Cauchy, Uniform, Gumbel). Note, the roles of R and S can also be exchanged.
  - The explicit formula for derivation of time-dependent reliability with consideration of survival is derived for scalar time effects only.
  - For the ability to transform the model to different reference periods (say, from fifty years to one year), S requires to follow an extreme value distribution (Fréchet, Weibull, Gumbel)
  - For more details, see Section 7.5.
- Visualization Techniques (Chapter 8)
  - Visualization by heat maps with respect to the stochastic distribution of a single variable of the reliability model is always applicable.
  - Creation of the presented visualizations is efficient when the considered stochastic variable is computed with SuSI or in the parameter state model by SuS.
  - The techniques can also be applied to simple models, but as SuSI and SuS perform best at complex structures, it is also most advisable to use these approaches for the latter.

Chapter 3

Subset Simulation

In this chapter, we review ordinary SuS (Au and Beck, 2001b), its state of the art, remark on benefits and drawbacks and discuss future opportunities and challenges. The reason is threefold. First, we want to give a better understanding of the algorithm, also to see how to increase its efficiency in Chapter 4. Second, the novel algorithm presented in this thesis (see Chapter 5) is based on Subset Simulation and therefore requires a substantial knowledge of the original algorithm. Lastly, also in Chapter 7, Subset Simulation is utilized as an optimal choice for efficient computation in the considered time-dependent reliability model.

3.1 State of the Art

3.1.1 The Original Algorithm

Subset Simulation (SuS) was introduced almost two decades ago by Au and Beck (Au and Beck (2001a), Au and Beck (2001b)). It is still a popular method in reliability evaluation for small failure probabilities, especially for high-dimensional problems. The main advantage might be its simplicity and general applicability without much prior knowledge about the problem at hand. It performs adaptively and works out of the box in most cases. SuS computes the failure probability of a general structural reliability formula, such as in (2.1), by computation of conditional probabilities. To do so, the fact that $g(\mathbf{x}) < b_i$ implies $g(\mathbf{x}) < b_{i+1}$ for $\mathbf{x} \in D$, $b_i, b_{i+1} \in \mathbb{R}$ and $b_i < b_{i+1}$ is utilized. A sequence of intermediate events $\mathbb{F}_1 \supseteq \mathbb{F}_2 \supseteq ... \supseteq \mathbb{F}_m, m \in \mathbb{N}$, is

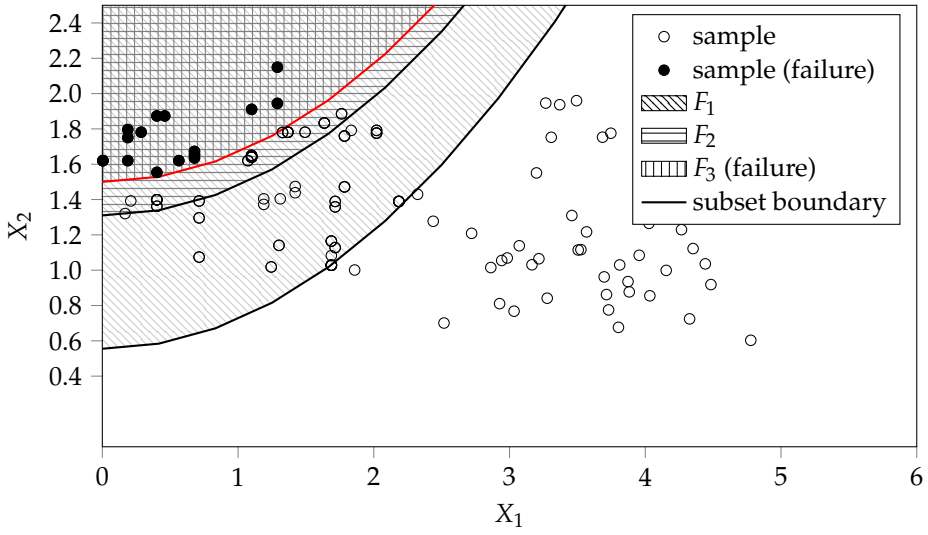


Figure 3.1: Blandfort et al. (2019a). Illustration of Subset Simulation with limit state function $g(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^2/3 - 2\mathbf{x}_2 + 3$, basic variables $\mathbf{X}_1 \sim \mathcal{N}(3.2, 0.92)$ and $\mathbf{X}_2 \sim \mathcal{N}(1.2, 0.13)$ and $m = 3$.

defined by $\mathbb{F}_i := \{\mathbf{x} \in D | g(\mathbf{x}) \leq b_i\}$ where $b_1 > b_2 > ... > b_m = b^*$ guarantees the subset property. By $b_m = b^*$ the last subset is associated with failure of the structure. We may thus compute the probability of failure p_f by

$$p_f = P(\mathbb{F}_m) = P\left(\bigcap_{i=1}^{m} \mathbb{F}_i\right) = P(\mathbb{F}_1) \prod_{i=2}^{m} P(\mathbb{F}_i | \mathbb{F}_{i-1}) \ .$$

This allows an effective evaluation of small failure probabilities by calculating higher conditional failure probabilities, which generally requires less computational effort. As a consequence, a smaller total number of samples E_T is necessary to achieve the same estimation accuracy as for classical Monte Carlo methods (Au and Beck (2001b)). The thresholds $b_1, ..., b_m$, defining the subsets, are chosen adaptively, after sampling a subset. One samples $\mathbb{F}_{i-1}$ in advance, then chooses a fraction $p_0 \in (0, 1)$ of samples with the lowest limit state values and defines the new threshold b_i as their highest limit state value. The conditional probability is then given as $P(\mathbb{F}_i | \mathbb{F}_{i-1})$ by definition of the threshold, where the corresponding estimate is given by p_0. We refer to this conditional probability as intermediate (subset) probability, also to avoid confusion with other conditional probabilities studied in this thesis.

Typically, one selects $p_0 \in [0.1, 0.3]$, which has been found to yield good efficiency (cf. Zuev et al. (2012)). When approaching a new subset, we only have $p_0 N$ samples distributed according to it, when proceeding this way. Since we need N samples again, a sampling procedure is necessary to create samples accordingly. Otherwise, the sample number would be continuously decreased, which would, most of the time, reduce efficiency significantly. The method of choice for creation of samples is the Markov Chain Monte Carlo (MCMC) method, which yields the main complexity of the algorithm but is often the key to success for analyzing unknown probability distributions. Typically a modified version of the Metropolis algorithm (Au and Beck (2001b)) or conditional sampling (Papaioannou et al. (2015)) are used. Since exploring properties when using MCMC has proved to be very difficult, the resulting algorithm is mostly treated like a Black Box (e.g. Seplveda and Faber (2019)) and has statistical properties which are hard to examine. For a pseudo-code of the algorithm, see Algorithm 1. Conventionally, seeds are kept (see Au and Wang (2014)) in SuS, only requiring to produce $(1 - p_0)N$ samples by MCMC in each subset. For detailed information, the reader is referred to the original work Au and Beck (2001b) and to Au and Wang (2014)[1]. In the next sections, we will also discuss state of the art in analysis and implementation of Subset Simulation and give a short overview on enhancements and improvements of the algorithm. For an easier discussion, we next introduce more notation regarding the SuS parameters. These have already been used in the given pseudo-code of the algorithm (Algorithm 1).

Definition 3.1.1 (Number of Samples, Number of Evaluations)**.** *We refer to the number of samples per subset by N. The resulting total number of samples is given by $N_T = m \cdot N$. The total number of evaluations is referred to as E_T and the corresponding number of evaluations per subset as E.*

The relation stated for the number of samples does not apply for measuring efficiency. Instead, the number of evaluations is the relevant characteristic. Further information is given in Section 4.1 and Section 5.2.

Definition 3.1.2 (Number of Subsets)**.** *The correct number of subsets in*

[1] A book containing a lot of details about Subset Simulation and the current state of the art. It also well demonstrates many opportunities for using Subset Simulation as the tool of choice for reliability evaluation, and provides a spreadsheet implementation.

Algorithm 1: Subset Simulation (SuS)

Result: Estimated Failure Probability $\hat{p}_f$

Initialization. Define
N (number of samples in each intermediate step),
p_0 (intermediate subset probability)

1. Procedure.
$i = 1$
create N Monte Carlo samples and evaluate limit state function g
threshold $b_i := p_0N$-th lowest limit state value of samples
while *subset threshold* $b_i > b^*$ **do**
  create N new or $(1 - p_0)N$ additional samples in subset $\mathbb{F}_i = \{\mathbf{x} : g(\mathbf{x}) \leq b_i\}$ by MCMC and evaluate them
  $i = i + 1$
  threshold $b_i := p_0N$-th lowest limit state value of samples
end
last subset level $m = i$
$\mathbb{F}_m = \{\mathbf{x} : g(\mathbf{x}) \leq b^*\}$
$\hat{p}_m = \frac{\#\text{samples in } \mathbb{F}_m}{\#\text{samples in } \mathbb{F}_{m-1}}$

2. Simulation Result.
$\hat{p}_f = \prod_{j=1}^{m} \hat{p}_j = p_0^{m-1} \cdot \hat{p}_m$

Subset Simulation with respect to p_0 and p_f, is given by

$$m = \left\lceil \frac{\log(p_f)}{\log(p_0)} \right\rceil .$$

We refer to the corresponding realization of the estimator of m by $\hat{m}$, reflecting the number of subsets for a specific termination of Subset Simulation.

Additionally, with respect to MCMC sampling, we specify specific characteristics that are always present in the sampling procedure.

Definition 3.1.3 (MCMC: Chain Elements)**.** *In each subset, the number of seeds is given by $N_s = p_0N \in \mathbb{N}$. The number of chains for MCMC is given by $N_c \in \mathbb{N}, N_c \leq N_s$ and the chain length of each chain is $N_l \in \mathbb{N}$.*

Observe that for $N_c = N_s$ we have to set $N_l = 1 + \frac{1-p_0}{p_0}$ to receive $N = p_0N + (1 - p_0)N = N_s + N_c \cdot (1 - N_l)$ subset samples. This choice is typical

in SuS implementations and only an appropriate definition for $p_0 \leq 0.5$. At first, those definitions might seem unnecessary with respect to the standard implementation where we always have $N_c = N_s$ and corresponding N_l. Later on, we will show how those can indeed become important and variation of those parameters changes efficiency of the algorithm drastically.

3.1.2 Enhancements: A Short Overview

Because of its high popularity, Subset Simulation has been significantly enhanced, improved and applied in many different fields of application, where many researcher have explored its strength and potential for analysis. Thereby, many ways to modify it for accelerated convergence or problem specific tasks were given. In the following, we give an extract of the developments of the algorithm and try to classify each according to its targeted benefits. We focus on topics relevant for, or closest to, our own contributions, the ordinary SuS algorithm, fundamental analysis and improvement thereof as well as modifications with the potential to efficiently analyze dynamic models.

Analysis and Enhancements of Ordinary SuS

Fundamental Analysis and Enhancements Subset Simulation was first presented in Au and Beck (2001b) (also compare Au and Beck (2001a)) to compute first-excursion failure probabilities of oscillating systems, e.g. a five-story shear building exposed to earthquake motion. Beyond the introduction of the novel algorithm, a fundamental analysis of its statistical properties was given. This analysis and the given discussion on challenges and assumptions, still define most of the basis for theoretical analysis of the algorithm to this day. In Zuev et al. (2012) the point estimate 'estimated failure probability', resulting by one run of SuS, was extended to a distribution of the true failure probability. The derivation is based on Bayesian statistics and the result is correspondingly introduced as 'Bayesian Post Processor'. Additionally to this, also a statistical analysis for selection of an optimal p_0 was given. Independent from the development in engineering, there was also progress in analysis of SuS in the rare event mathematical literature. There, however, the procedure is named 'Sequential Monte Carlo' or 'Splitting' and similar algorithms were first introduced in Cérou et al. (2006) and Johansen et al. (2005). In particular, Cérou et al. (2012) is a good reference for asymptotically derived results on statistical analysis. A very detailed overview on

several aspects regarding SuS is given in Au and Wang (2014) that also sums up many previous findings and discusses some open questions. The book possibly makes the best basis for introduction to the topic 'Subset Simulation' and also yields a spreadsheet implementation of the algorithm. Papaioannou et al. (2015) and Au and Patelli (2016) focus on efficient MCMC sampling and allow to decrease the dependency of generated MCMC samples in SuS, providing increased efficiency of the algorithm. More information on these contributions is given in Section 3.1.3, where some results are explicitly stated and discussed in more detail.

Increased Efficiency Trough Modified or Extended Approach Besides fundamental analysis and improvements of the algorithm, there has also been development on modifications of the basis of Subset Simulation. For example Ching et al. (2005) utilizes splitting instead of MCMC to improve the efficiency for reliability estimation of dynamical systems, applicable to first-passage problems. In Katafygiotis and Cheung (2007), spherical Subset Simulation which computes the failure probability by dividing the failure domain into subregions and the auxiliary domain method, are presented. Researchers keep the algorithm under constant development up to date. Recently it was proposed in Abdollahi et al. (2020) to use a fitness-based seed selection in SuS, preferring more probable samples over others in MCMC for increased efficiency.

In contrast to beforehand mentioned changes of the core of SuS, many approaches that aim at an increased efficiency by utilizing meta models have been developed. For example Bourinet et al. (2011) uses support vector machines and in Papadopoulos et al. (2012) neural networks are utilized to increase efficiency of SuS. This well demonstrates that there is still much room for enhancement of SuS, as the algorithm is very generic and provides non-failure and failure samples whose implicit information can be used in many ways.

Approaches with a Potential for Efficient Evaluation of Dynamic Models

An important field in research is the modification of the algorithm to suit demands of a specific selected engineering problem. By adapting SuS to a given problem type, which can then be calculated more efficiently, some generality is lost and other cases are solved less efficiently. Nevertheless, such

approaches are important to solve certain problems in a feasible time or to generally achieve the maximum performance possible. Some of the developed methods can potentially be used for dynamic model evaluation. This paragraph aims at providing a brief (and certainly not exhaustive) overview on such methods, to clarify the difference of existing approaches to our novel algorithm, presented in Chapter 5.

In Au (2005), an augmented reliability problem is considered, allowing to use SuS for efficient calculation of design sensitivity. The paper focuses on design optimization and is based on selecting at least one originally deterministic design parameter to be stochastic. Using ordinary SuS on this augmented reliability problem then provides information on reliability with respect to selection of this design parameter. Such conditional failure probabilities are then presented by means of a histogram, where larger intervals reduce the introduced stochastic noise. Note in particular that Au (2005) does not provide failure probabilities with respect to distributions of the selected parameter as it was found that spurious noise will typically be introduced on the probability density function. Such an extension would most likely need a high number of samples to get a good estimate, particularly in the subsets that contain much information on failure probabilities. Song et al. (2009) focuses on sensitivity of the failure probability with respect to the distribution parameters of the stochastic variables. This method provides local information. To calculate the reliability of several limit state functions, Hsu and Ching (2010) derives one variable correlated with all performance functions by a principal component analysis and then conditions on this representative to evaluate approximate failure probabilities. If multiple stochastic responses need to be considered, generalized Subset Simulation (Li et al. (2015a)) can be a good choice. It tackles the 'sorting difficulty' for choosing subset seeds when multiple limit states are considered so that sorting according to the limit state value is not unique anymore. It constructs a unified intermediate event instead, allowing to evaluate several limit state functions simultaneously. Boundaries for failure probabilities, using random set theory, are computed in Alvarez et al. (2018). A sequential approach for time-dependent analysis of deteriorating structures is given in Kim and Straub (2019) (also compare Schneider et al. (2017)). There, it is utilized that the reliability of deteriorating structures decreases with respect to time and thus allows to reach the failure region by SuS faster when starting with the last relevant time instant. Starting in reverse order then allows to evaluate failure probabilities with respect to the limit states at different times

in an efficient way, using samples of previously evaluated time instants for successive evaluation of all time instants.

Subset Simulation Interpolation (SuSI): A Different Approach

The discussed methods focus either on local sensitivities or on evaluation of a selected number of constellations. If many limit states are considered and particularly a dynamic variable $\mathbf{X}_k$ can take substantially differing distributions, such methods may result in high computational efforts. In contrast, when using SuSI, evaluation of several constellations needs no additional computations so that we are not so much affected by the number of considered limit states with respect to the selected distribution of the dynamic variable $\mathbf{X}_k$. To achieve this, our presented algorithm has two new main features:

- The way subsets are created is modified substantially, evaluating failure sets with respect to the dynamic model variable $\mathbf{X}_k$. This allows to collect information on failure regions in many (up to all) subsets, exploring the sample space very efficiently.

- A conditional failure probability function, allowing to evaluate reliability with respect to any distribution of $\mathbf{X}_k$, is provided by extension of the discrete solution set of the individual subset levels by interpolation.

With respect to the first point, note that developing meta models based on this information could provide an even more efficient algorithm. Additionally, a new way to explore the sample space for reliability evaluation by MCMC provides a completely different way to derive a solution for a reliability problem while keeping the general remarkable properties of SuS. This might not appear to be a great benefit at first, but it does, among others, provide the opportunity for increased evidence on results by calculation solutions with several methodologies while having the big advantage that SuSI and SuS perform well on the same type of problems, not needing to account for a not well suitable method for verification of results. As a consequence of the second point, high accuracies for many constellations of $\mathbf{X}_k$ can be achieved at the costs of a small number of total evaluations.

Remark 3.1.4. *Although based on Subset Simulation and keeping most of its remarkable properties, there is no direct competition to most existing methods*

based on Subset Simulation since such modifications might additionally be utilized in the presented approach. Instead, its novelty guarantees interesting fields of study based on the provided knowledge of enhancements of Subset Simulation.

The fundamental difference of the novel algorithm makes it particularly interesting to explore opportunities for combination with existing methods such as Au (2005) or Song et al. (2009), as the samples in individual subsets have not been utilized further to its fullest potential yet, as in some of the discussed methods, in our algorithm. We could possibly explore uncertainty with respect to several dimensions at computational costs similar to a static reliability evaluation with SuS. This could deliver important information in such settings, because, as stated in Au (2005), the computational demands typically grow exponentially with the number of variables to be explored, which generally seems to be an inevitable drawback when targeting on multi-dimensional information and prevents us from such a sophisticated analysis.

Besides the advantages of the algorithm, there are of course also disadvantages as well. First, our dynamic model variable is restricted to the assumptions it is based on, monotonicity with respect to the limit state function and independence of the dynamic variable with respect to all other variables of the model (compare Section 2.2). However, note that the dynamic variable does not need to be a variable of the original problem formulation and can be artificially created instead. Second, although the costs are similar as in SuS, they are often slightly higher because some extra evaluations are necessary in the new way subsets are created. As last and possibly worst disadvantage, SuSI has a higher complexity in analysis and implementation compared to SuS.

3.1.3 Brief Summary of Assumptions and Selected Fundamental Analytical Results

In this section, we outline the fundamental state of the art in analysis of the algorithm, to start with a strong basis for discussing our own contributions later.

Two very important points that define the character of the algorithm "Subset Simulation" are estimation by conditional probabilities and sample creation by Markov Chain Monte Carlo. Because Subset Simulation samples from an unknown probability distribution and tries to find certain regions of

the sample space that induce failure by the available information by already created samples, it can be understood as an adaptive sampling method or similarly as an importance sampling scheme where the importance sampling function is implicitly computed adaptively on the fly. The algorithm was first introduced in Au and Beck (2001b), already providing a strong fundamental statistical analysis of SuS and defining the basis for analytical results on SuS up to date. The general procedure for application is also shortly discussed in Section 3.1.1. In the following, we will formally account for extensions in understanding of the algorithm by analysis as well as discuss steps towards state of the art implementations with highest efficiency available today. In detail, we briefly review the topics 'Assumptions and Basic Properties of SuS', 'Efficient MCMC Sampling', 'Analytical Results: Bias and Variance Formula' and 'Optimal Intermediate Probability p_0 Selection'. We start by listing relevant assumptions, made to draw conclusions in analysis.

Assumptions and Basic Properties of SuS

This paragraph gives an overview on some assumptions made in many approaches for analysis of SuS. All these assumptions were first introduced in Au and Beck (2001b).

An often made assumption that mainly focuses on a simplification of notation is the following.

Assumption 3.1.5 (Integer Number of Subsets)**.** *The number of subsets m is assumed to be integer*

$$m = \frac{\log(p_f)}{\log(p_0)} \ .$$

In general this does not result in much loss of generality. Another assumption that, on the other hand, is crucial and restrictive at the same time is that N is big enough.

Assumption 3.1.6 (Asymptotic Considerations)**.** *We assume high sample numbers N.*

Our results are asymptotic ones because it is not clear how high N needs to be for the, under this assumption, drawn conclusions. However, moderately high N (e.g. $N = 1000$ for $p_0 = 0.1$) are expected to lead to sufficiently accurate results, in line with theoretical derivations, often. Assumption 3.1.6 is very important to proceed with an appropriate statistical analysis in many

cases. Another assumption, often made, is an a priori selection of the subset thresholds.

Assumption 3.1.7 (Pre-defined Thresholds)**.** *We assume the subsets $\mathbb{F}_i$, $i = 1, ..., m$, or equivalently the thresholds b_i to be defined a priori so that the true intermediate probabilities are equal to p_0 in every subset.*

This allows to consider the estimated intermediate probabilities in SuS as stochastic, which they indeed are not in real implementations of the algorithm where they are set to the constant p_0 for all subsets except for the last one. Still, results derived under such an assumption should be similar for high N, and hence often lead to good enough results while they may simplify analysis drastically. Thus we will stick to Assumption 3.1.7 in the introduction and analysis of our new algorithm Subset Simulation Interpolation. On the other hand, since Chapter 4 is dedicated to a rigorous stochastic analysis of the properties of Subset Simulation, we will drop it in the following chapter in order to derive exact results. Next, regarding MCMC sampling, we can not guarantee the samples to be distributed according to the true distribution in general for small or moderately high N. We refer to such considerations under the term 'ergodicity'.

Definition 3.1.8 (Ergodicity)**.** *MCMC sampling or its corresponding MCMC chains are called ergodic, if the distribution of samples created by MCMC sampling follow the limiting stationary distribution.*

In Subset Simulation, we sample according to the true distribution of the corresponding subsets if ergodicity is given.

Assumption 3.1.9 (Ergodicity)**.** *The Markov chains in Subset Simulation are assumed to be ergodic, i.e. their distribution does not depend on the initial state.*

For high N, the algorithm becomes more and more similar to a robust standard Monte Carlo simulation so that for high enough N we will have ergodicity. Often used in statistical analysis at first, followed by a later adjustment to real settings, is the following assumption that would hold under perfect MCMC sampling.

Assumption 3.1.10 (Within Subset Sample Independence)**.** *The samples in each subset are considered mutually independent.*

Efficient MCMC Sampling

First of all, note that we do not focus on MCMC. Thus, we will just discuss it very shortly, although it is a crucially important topic regarding efficient Subset Simulation. For more detailed information, than provided here, on this topic, we refer to Au and Wang (2014), Papaioannou et al. (2015), Au and Patelli (2016).

For MCMC sampling, a modified version of the Metropolis algorithm was introduced in Au and Beck (2001b), which allows for a much more efficient sampling in high dimensions. Choosing the proposal pdf[2], the distribution according to which new potential samples are created in MCMC sampling, was already identified as challenging. This distribution defines how new subset samples are produced by the subset seeds. The most critical part is the selection of an appropriate 'distance' from the last chain element (by setting the variance of the proposal pdf) where new samples are created preferentially. It requires finding a good balance between not rejecting too many samples as a consequence of a too high variance of the proposal pdf and sampling very close to original samples due to a low variance. Too high or low variance both result in highly correlated samples by MCMC.

Remark 3.1.11. *In traditional SuS implementations (e.g. Au and Beck (2001b)), the variance of the proposal pdf is kept constant over all subsets, which we will later refer to as non-adaptive methods.*

The modified Metropolis algorithm works well for high dimensions and fulfills the so-called reversibility condition[2], which guarantees that samples produced by MCMC remain in the same distribution as the original sample they are created from.

Remark 3.1.12 (Reversibility Condition)**.** *All considered MCMC sampling procedures in this dissertation satisfy the reversibility condition.*

Thereby, if the original samples, used as seeds for a Markov chain, are in the stationary distribution, new created samples by MCMC are also in the stationary distribution in all considered MCMC sampling procedures. More information and the corresponding proofs regarding the reversibility condition are given in Zuev et al. (2012), Au and Wang (2014) and Papaioannou et al. (2015).

[2]See e.g. Au and Wang (2014) or Papaioannou et al. (2015) for more details.

Regarding the challenging selection of the proposal pdf, current implementations provide an adaptive procedure. This procedure can improve efficiency by orders of magnitude and also substitute the modified Metropolis algorithm by an efficient conditional sampling procedure (Papaioannou et al. (2015), Au and Patelli (2016)).

Remark 3.1.13. *In modern SuS implementations (see Papaioannou et al. (2015)), the proposal spread for sampling is modified on the fly, so that the acceptance rate of the samples stays close to an a priori selected target value, known to provide good efficiency of the MCMC sampling.*

A discussion on the relation of efficiency and acceptance rate in MCMC sampling is also given in Zuev et al. (2012).

By the reversibility condition, it is known how to remain in an attained stationary distribution belonging to a subset of Subset Simulation and that all considered approaches in this dissertation do so (Remark 3.1.12). Whether the sample seeds by MCMC are in the stationary distribution from start or not is thus the next important point to be discussed and is closely connected to ergodicity (Definition 3.1.8).

Remark 3.1.14 (Ergodicity: Number of Chains and Proposal Pdf)**.** *When the number of MCMC chains or the variance of the proposal pdf is small, the assumption of ergodicity (Assumption 3.1.9) can be violated. An easy example for such a violation of ergodicity is given if there is only one MCMC chain created and there exist several unconnected failure regions which are separated by regions which too large to be crossed by the MCMC steps (compare Au and Beck (2001b)). Then, one chain alone can not explore all relevant regions at the same time so that the initial state of the Markov chain becomes relevant, i.e. we are not in the stationary distribution of the simulated subset. However, if the variance of the proposal pdf is chosen high enough, we may always achieve ergodicity (Au and Beck (2001b)).*

This concern regarding ergodicity is, among others, discussed in Au and Beck (2001b). However, for a statistical analysis, we have to assume that ergodicity holds, since otherwise it is not clear at all which distribution we sample from. The consequence is Assumption 3.1.9. A high variance of the proposal pdf makes fulfillment of the condition more likely. Furthermore, it is noted in Au and Beck (2001b) that the subset seeds already follow the correct distribution by construction, as stated in the following remark.

Remark 3.1.15 (No Burn-In)**.** *The seed samples landing in intermediate subsets are distributed according to the corresponding subset. Correspondingly, the stationary distribution in MCMC is reached immediately.*

Thus, no burn in, which is a typical drawback in MCMC sampling, is required. These points taken together, having ergodicity and keeping the correct distribution by MCMC sampling, means that we are indeed sampling from the desired distribution of the subsets.

We can only sample from this specific distributions at the cost of dependency of the created samples by MCMC sampling. Therefore, it is recommended to compute the effective independent subset sample number in each subset $\mathbb{F}_i$ after termination of the algorithm.

Assumption 3.1.16 (Effective Number of Samples)**.** *The effective number of samples in subset $\mathbb{F}_i$ with N samples is assumed to be given by $N/(1+\gamma_i)$ with (Au and Beck (2001b))*

$$\gamma_i = 2 \sum_{k=1}^{N/N_c-1} \left(1 - \frac{kN_c}{N}\right) \rho_i(k) \ ,$$

where $\rho_i(k)$ is the empirical autocorrelation of lag k in subset i.

The autocorrelation corresponds to the sequence of indicator variables of an MCMC chain, indicating whether the sample is in the next subset or not $\{I_{j,k}^{(i)} : k = 1, ..., N/N_c\}$. For an exact derivation, see Au and Beck (2001b). We refer to γ_i as the correlation factor of subset i, $i = 1, ..., m$ and $\bar{\gamma}$ is defined as the average correlation factor over all subsets levels. Assumption 3.1.16 implicitly assumes that different MCMC chains are uncorrelated with respect to the indicator function. A comprehensive overview, covering these and some additional aspects of MCMC in SuS is also given in Au and Wang (2014). Under usage of the correlation factor, we can now reformulate Assumption 3.1.16 as setting $\gamma_i = 0$ for all $i = 1, ..., m$. In general, an average dependency term $\bar{\gamma}$ is often first left out of analysis and added afterwards. This simplifies analysis considerably.

Analytical Results: Formulas for Bias and Variance

All derived formulas require Assumption 3.1.9, meaning that MCMC sampling generates samples of the desired distribution.

First, the coefficient of variation (CV) of each intermediate subset probability estimator is considered locally. Under Assumption 3.1.16, the CV of the Bernoulli distributed probability estimator for the intermediate probability $p_0 = P(\mathbb{F}_i|\mathbb{F}_{i-1})$, $i = 1, ..., m$, is given by

$$\delta_i = \sqrt{\frac{1-p_0}{p_0 N}(1+\gamma_i)} \ .$$

This local result is then utilized to derive results for estimation by SuS. Under assumption of pre-defined thresholds (Assumption 3.1.7) and high N (Assumption 3.1.6) the intermediate subset probability estimators naturally follow a Gaussian distribution with known mean value p_0, by the Central Limit Theorem. This allows to derive boundaries for the relative bias (writing B for the absolute bias) of the failure probability estimator $\hat{P}_f$ with respect to the true failure probability p_f as (Au and Beck (2001b))

$$\begin{aligned}\frac{|\mathrm{B}|}{p_f} &= \left| E\left[\frac{\hat{P}_f - p_f}{p_f}\right]\right| \\ &\leq \sum_{i>j} \delta_i \delta_j + o(1/N) \\ &= O(1/N)\end{aligned}$$

and coefficient of variation (CV) of $\hat{P}_f$ given by (Au and Beck (2001b))

$$\begin{aligned}\mathrm{CV}^2 &= E\left[\left(\frac{\hat{P}_f - p_f}{p_f}\right)^2\right] \\ &\leq \sum_{i,j=1}^{m} \delta_i \delta_j + o(1/N) \\ &= O(1/N) \ .\end{aligned}$$

Note, this corresponds to a coefficient of variation that is $O(1/\sqrt{N})$. Under assumption of uncorrelated intermediate subset probability estimators, this reduces to

$$\mathrm{CV}^2 = \sum_{i=1}^{m} \delta_i^2$$

which was found to well approximate results in simulations (Au and Beck (2001b)). Also including the possible correlations between intermediate subset probability estimators, another formula (under reformulation) is given in Au and Beck (2001b):

$$\mathrm{CV} = \sqrt{|\log(p_f)|^r \frac{(1+\bar{\gamma})(1-p_0)}{p_0 |\log(p_0)|^r N_T}} \ , \tag{3.1}$$

with $r \leq 3$ considering the true correlation of the intermediate subset probability estimators.

In contrast to this analysis, in the mathematical literature, analysis of the properties of SuS was done also by order statistics (Cérou et al. (2012)). Here, it is assumed that the number of subsets coincides with the correct number thereof (compare Definition 3.1.2), which is fulfilled with probability one for $N \to \infty$ under weak conditions. Nevertheless, having e.g. $p_f = 10^{-4}$ and $p_0 = 0.1$ violates this assumption, since then asymptotic arguments fail as remarked in Cérou et al. (2012). Anyhow, asides from that, not much generality is lost. The proof relies on derivation of the distribution of the threshold of the last subset, because this threshold alone then yields the estimated failure probability. For deriving an approximation for bias and variance, a Taylor approximation is utilized. Although this is a completely different approach, the resulting formula for the CV is similar, but does not consider dependencies within subset samples, yielding (Cérou et al. (2012))

$$\mathrm{CV} = \frac{1}{\sqrt{N}} \sqrt{m \frac{1-p_0}{p_0} + \frac{1-p_m}{p_m}} \ . \tag{3.2}$$

Additionally, an asymptotic formula for the bias is given (Cérou et al. (2012))

$$E\left[\frac{\hat{P}_f - p_f}{p_f}\right] \underset{N \to +\infty}{\longrightarrow} \frac{1}{N} \frac{m(1-p_0)}{p_0} \ .$$

where the formula was modified according to Assumption 3.1.5. Furthermore, the distribution of the estimator is shown to be asymptotically normal, yielding a complete definition for the distribution of the Subset Simulation result. The requirements for this analysis are assumption of independence within subset samples (Assumption 3.1.10) and high sample numbers N (Assumption 3.1.6). Also the counterpart of Assumption 3.1.5 is necessary. Note

that Equation 3.2 coincides with Equation 3.1 under $\bar{\gamma} = 0$ (independent within subset samples), $r = 2$ (independent intermediate subset probability estimators) and Assumption 3.1.5 (last intermediate probability is p_0), when substituting $N = \frac{N_T}{m}$ and $m = \frac{\log(p_f)}{p_0}$. Furthermore, for SuS under fixed thresholds (Assumption 3.1.7), unbiasedness is claimed in Cérou et al. (2012).

A third approach for analysis of the statistical properties of SuS is followed in Zuev et al. (2012). In contrast to the other approaches, no asymptotic arguments (no need for Assumption 3.1.6) are necessary and $\hat{p}_f$ is assumed to be given and p_f is analyzed as if it was stochastic, corresponding to a Bayesian approach after given a terminated SuS simulation run. The focus of this analysis is the derivation of a Bayesian Post processor, giving a distribution of an estimator for the true failure probability P_f after termination of SuS instead of receiving only the point estimate $\hat{p}_f$. Thereby, confidence intervals for the true failure probability may be derived. A necessary assumption for derivation of the post processor is independence of within subset samples (Assumption 3.1.10). This assumption is necessary to derive closed formulas most of the time but unfortunately contradicts the creation of samples by MCMC, as utilized in Subset Simulation. More information is given in Section 4.2, as our approach for analysis is similar, but differs in some substantial points such as in the assumed distribution of the intermediate subset probability estimators and threshold selection. In Zuev et al. (2012), intermediate probabilities are also assumed Beta distributed as in the presented approach. Note, we use different parameters since our analysis is based on order statistics.

Optimal Intermediate Probability $\mathbf{p_0}$

At first, recommendations for selection of p_0 were made on the basis of simulations, where $p_0 = 0.1$ yields good results (compare Au and Beck (2001b)). Later, a statistical derivation was given in Zuev et al. (2012), minimizing the coefficient of variation with respect to p_0. Utilized assumptions are, beyond using an asymptotic formula (Assumption 3.1.6), independence of the average within subset sample dependency $\bar{\gamma}$ of the selected intermediate subset probability p_0 and that no seed samples are re-used. It is concluded that the coefficient of variation is insensitive with respect to p_0 for $p_0 \in [0.1, 0.3]$, reaches its optimum at $p_0 = 0.2$ and becomes very bad for exceptionally small or high values of p_0. This is the state of the art recommendation for

an optimal choice of p_0 up to date. Under re-use of seeds (there is consensus to do so), the formula however yields flat results for high p_0 also. This unfortunately is found to contrast results by simulation, where the coefficient of variation appears to increase for high p_0. In Au and Wang (2014), this finding is justified by an increased correlation between intermediate subset probability estimators for increased values of p_0.

3.2 Challenges

In this section, we state challenges in analysis and application of SuS. Indeed there are many assumptions that have to be made for an adequate statistical analysis and a few of those are clearly violated in some cases. This might not always result in too bad consequences, but nevertheless it remains important to discuss those and look for improvement. Most points have already been identified in past research, but have not been solved so far. If we provide a (sometimes only partial) solution to the stated challenge, a reference to the corresponding section is given.

Asymptotic Analysis and Reality (Section 4.2). An important aspect which is particularly interesting with respect to derivation of practical implications by statistical analysis is given in the next remark.

Remark 3.2.1. *[Small Sample Numbers] As Subset Simulation focuses on a reduction of computational demands and should be applied by repetitive runs of the algorithm, small sample numbers (e.g. $N = 1000$) or even very small sample numbers (e.g. $N = 200$) are of interest. The following three main reasons for the relevance of small sample numbers are identified.*

- *Most importantly, one often uses trivial parallelization by splitting the independent Monte Carlo runs into smaller chunks. Evaluation of a reliability problem by Subset Simulation is typically performed by several independent runs of the algorithm and a subsequent averaging over all results. This may result in a small sample number of samples in each SuS evaluation, even when a higher number of total evaluations is feasible. Benefits by repetitive independent runs of the algorithm are possible feedback on ergodicity considerations and achievement of an empiric, problem specific, coefficient of variation. Both points are important for a reliable assessment of the result by SuS.*

- *Very expensive evaluations only allow for small N. This even becomes increasingly relevant due to an ongoing increasing model complexity in the engineering sciences.*

- *High intermediate subset probabilities p_0 can result in small N. For any given constant number of total sample numbers N_T, the sample number per subset N can be arbitrarily decreased by increasing p_0.*

In contrast to Remark 3.2.1, statistical analysis is often based on asymptotic arguments, made under the assumption of high sample numbers N per subset. Results are then extrapolated to the lower sample cases, where accuracy remains unclear (see Breitung (2018)). This is especially the case for derivation of bias and coefficient of variation that aim at optimization of the algorithm. In addition, asymptotic results avoid strict comparisons of theoretical claims and verification of those by simulation, since discrepancies might be caused by approximations. As a consequence, it is hard to judge whether new manipulations of the algorithm improve it or not. Simulation studies are computational demanding and only cover specific cases. The Black Box 'Subset Simulation', typically applied on many substantially differing applications as discussed in Chapter 1, is thus challenging to interpret and to alter in a meaningful way.

Ergodicity, Stationarity and Adaptive Sampling (Section 6.2). If the failure regions are not connected or if the steepest descent of the limit state function, taken in the SuS procedure, does not lead to the most relevant failure regions, then we might not get the correct results if the proposal spread is not chosen exceptionally high. Several examples are given in Breitung (2019), where it is shown how specific geometries of the limit state function lead to bad exploration of the sample space and thus highly biased SuS results. For higher N and high proposal spread, this issue will vanish. However, then the demands might be almost as high as for crude Monte Carlo. In the new adaptive methods, desired acceptance probabilities in MCMC sampling are tried to be kept also in higher subset levels. Accordingly, the proposal spread is successively decreased and therefore ergodicity issues become even more likely for higher subset levels than in the standard procedure. The performance however is much better in the new methods so that we still prefer these over the non-adaptive ones.

Remark 3.2.2 (Ergodicity: Geometry of Limit State Function)**.** *There are cases, where SuS does not explore the state space appropriately, leading to systematically wrong results for typical choices of N. Some examples, resulting in a (sometimes heavily) biased result have been given in Breitung (2019) and Bourinet (2018).*

In particular, such systematically wrong results are typically (dangerous) underestimations of the failure probability. It is thus extremely important to find criteria for circumventing this problem. A promising approach is presented in Section 6.2.

Stationary Distribution and Thresholds (Section 4.2). Another critical point is the way subsets are created. Indeed, the threshold of a subset is selected so that it belongs to the limit state value of a sample seed, which is contained in the same subset. Under re-use of seeds, this sample will surely be contained in the next subset. If seeds are not re-used, still the new samples are correlated with the given sample. This sample however does not follow the stationary distribution of the next subset as discussed in Breitung (2018). On the one hand this will typically not have a significant impact on the result. For low N, on the other hand, it could.

Remark 3.2.3 (Threshold Selection and Bias)**.** *The sample seed which defines the threshold of a subset produces a bias on the SuS result.*

So, in ordinary SuS there is a, to the authors best knowledge, not yet analyzed bias.

Dependency or Independence of Intermediate Subset Estimates (Section 4.2). The dependency structure between intermediate subset probability estimators is unknown. Knowing the real dependencies would be substantial for statistical analysis. It was found in simulations, that statistically derived formula for the coefficient of variation yield good results for low p_0, when independence is assumed. In contrast to that, for high p_0 it is recommended to assume increasing dependencies between intermediate subset probability estimators Au and Wang (2014). In Section 4.2, we show independence of successive intermediate subset probabilities under specific conditions on the sampling.

Many Existing Implementations (Section 4.1). There exist many different implementations for SuS. As a consequence, it is hard to generally draw conclusions as those might only be valid for specific implementations and not suit others at all. This makes it important to specify the implementation, an analytical result is based on.

Inconsistency of Statistically Derived CV and Results by Simulation and Thus Optimal Intermediate Probability Selection (Section 4.1). The statistically derived CV formula fails to provide even approximate results for high p_0. The discrepancy is big and therefore indicates serious differences between theory and practice. A possible reason was given by an increased correlation between intermediate subset probability estimators when higher p_0 are considered (see e.g. Au and Wang (2014)). Even discrepancies to the existing recommendation of $p_0 \in [0.1, 0.3]$ by statistical analysis (Zuev et al. (2012)) and performance in simulations have been found at some points, such as in Li and Cao (2016). It was found that $p_0 = 0.3$ leads to a significantly worse performance than $p_0 = 0.15, 0.2, 0.25$. Note that inconsistencies in CV formula and results by simulation do not allow to adequately select an optimal p_0 based on a theoretical basis, having a direct impact on efficiency in application of SuS.

Distribution of the Result (Section 4.2). In general, the distribution of the estimator by SuS is unknown. For high N it is Gaussian by the Central Limit Theorem. However, for low N it was shown to rather look like a lognormal distribution (Breitung (2019)). As a consequence of the inconsistent or unknown distribution, it is of course also difficult to derive formulas for bias and variance of the result. Also the derivation of good confidence intervals for a small number of SuS realizations in practical application can be challenging. We propose to split statistical analysis in high and low sample numbers generally, because of the crucial differences in both the cases and particularly since many considerations are only relevant for low sample numbers (compare Chapter 4). In particular formulas for the coefficient of variation can not hold for both cases since for low sample numbers the heavy tailedness of the lognormal distribution will drastically increase the CV in comparison to higher sample numbers due to the difference in the shape of the distributions.

Repetitive Runs for Evidence by Simulation. In particular for small sample numbers, the distribution of the result of SuS becomes increasingly heavy-tailed. Examining such distributions by repetitive runs in a simulation study needs exceptionally many simulation runs if exact results are required. A few samples will lead to extremely high values and contribute significantly to the total result. However, for higher sample numbers repetitive runs become increasingly computational demanding, so that the number of repetitive runs has to be decreased often. It is thus challenging to use simulation studies for reliable results. If several independent studies are made, less independent runs might be sufficient. Then, however, one has to be aware of the variation of such results and should not consider them as precise outcomes.

Remark 3.2.4 (Necessity of Many Simulation Runs: Slow Convergence)**.** *Simulation studies with Subset Simulation require many independent runs to draw valid conclusions. The distribution of its results is given by a product of random variables, resulting in a heavy tailed distribution of the results for many N. If simulations should deliver exact results, e.g. when specific minor effects have to be explored, many simulations are necessary to receive trustworthy results. A good choice would be to create at least* $10,000$ *independent simulation runs. This is a recommendation for testing one constellation with respect to theoretical analysis where exact results are required. If several similar constellations are tested and only trends or less sensitive characteristics are explored, then also fewer (e.g.* $500 - 2000$*) repeated runs should suffice the demands on accuracy. Furthermore, the feasibility of computation may set lower limits.*

Concluding, depending on the conclusion discussed, the number of repetitive runs for study has to be adapted carefully.

Consequence: Theoretical Study and Simulation. As a consequence of the small sample numbers in simulations, the threshold selection, the distribution and the unknown dependencies between intermediate subset probability estimators, it is hard to predict effects of changing parameters in SuS or to alter the algorithm in a meaningful way. There is evidence in simulations on claims by statistical analysis, but there are also high discrepancies in some cases. An example is the increasingly high CV for high p_0 values. The question is, whether such discrepancies are the consequence of small sample numbers, wrong assumptions or just a matter of approximation

and simplification in statistical analysis in general. Explanation of such discrepancies then needs to be assigned to specific shortcomings of the analysis such as e.g. dependent intermediate subset probability estimators, the use of small sample numbers or a low number of repetitive independent simulation runs for examination. If the drawn conclusions suit reality, can often not be shown with confidence. Chapter 4 partially bridges the gap between theoretical study and practical findings in simulations, analyzing SuS in detail, particularly with respect to the above presented challenges.

Chapter 4

Subset Simulation: Enhancements and Extended Stochastic Analysis

As summarized in Chapter 3, there are many variants of ordinary Subset Simulation. Aiming at general conclusions for Subset Simulation in this chapter, we still need to differ between implementations, but only require to consider properties which have to be specified in every implementation of Subset Simulation. First, we always need to decide whether seeds of previous subsets are re-used for evaluation of successive subsets or not. Second, the type of Markov Chain Monte Carlo (MCMC) sampling is chosen to be either non-adaptive or adaptive. Adaptiveness refers to adaption of the spread of the proposal distribution to increase efficiency of the sampling (cf. Remark 3.1.13). A difference between conditional sampling and modified Metropolis algorithm is not made due to the comparably little difference in efficiency for high-dimensional problems (compare Papaioannou et al. (2015)).

Definition 4.0.1 (Implementations)**.** *Subset Simulation is naturally implemented according to one of the methods (I1)-(I4), given in Table 4.1.*

(I1) and (I2) refer to the traditional sampling such as in Au and Beck (2001b) where (I3) and (I4) represent modern implementations. For more details on (especially modern) MCMC methods, see Papaioannou et al. (2015). In addition to the existing methods (I1)-(I4), we also study additional, though also essential, variants of the most efficient implementation (I4), given by (I5) and (I6). Those have not been examined yet. A detailed introduction

Established methods		
Method	Re-Using Seeds	Adaptive MCMC
(I1)	No	No
(I2)	Yes	No
(I3)	No	Yes
(I4)	Yes	Yes
New methods, based on (I4)		
Method	Bias Correction	Drop Threshold Seeds
(I5)	No	Yes
(I6)	Yes	Yes
* If a star is added (e.g. (I4*)), we use the fixed chain length $N_l = 10$		

Table 4.1: The Basis of Subset Simulation Implementation: We only consider differences in inevitable characteristics of the implementation of the Subset Simulation algorithm. These decisions must be made in every implementation. Usage of bias correction and threshold seeds are additional decisions that were not examined before in the engineering literature, never using a bias correction and always using threshold seeds in (I1)-(I4) by default.

is given in Section 4.2. As recommended in Section 3.2, we split analysis of SuS into moderate to high and small sample numbers. Although most results presented hold for both cases, there are good reasons to do so. We next state two of them. First, although the claims that will be made in Section 4.1 appear to be also valid for small N, the utilized formula for the coefficient of variation was originally derived under the assumption of high N. Second, the conclusions in Section 4.2 are more important for small sample numbers because the studied effects will often, except for independence of intermediate subset probabilities (Remark 4.2.17), have a negligible impact on results for high sample numbers. Our practical simulations are carried out under adaptive conditional sampling with re-use of seeds (I4), according to Papaioannou et al. (2015), and later with (I4*), (I5*) and (I6*). So, although non-adaptive methods seem to be popular still, we will only use the most efficient implementations for our simulations, since these offer the most relevant information for future research and application.

4.1 Optimal Choice of Intermediate Subset Probabilities for Moderate and High Sample Numbers

Outline

The optimal choice of the intermediate subset probability p_0, has not been derived under explicit differentiation of available implementations (I1)-(I4) yet. Therefore, we extend the existing analysis, showing that under current advances, the common choice of $p_0 = 0.1$ (Au and Beck (2001b) or $p_0 \in [0.1, 0.3]$ (Zuev et al. (2012)) is not necessarily optimal anymore. Instead, the length of the Markov chains in Subset Simulation is identified as the most important parameter for efficiency. In addition, we extend the analysis to support choices such as $p_0 = 0.2$ as an optimum in traditional implementations and for improvement of the general understanding of Subset Simulation and discuss an aspect regarding the within-subset-sample correlations with respect to the threshold selection. Our results in this section will often be based on asymptotic arguments, as we condition on an asymptotically derived formula for the coefficient of variation. We thus recommend to relate the derived results to moderate or high sample numbers and not necessarily for very small sample numbers where we have to additionally account for the results of Section 4.2.

Preliminaries

Our objective is to choose the intermediate subset probability p_0 so that we achieve the best possible accuracy (measured by CV) of the Subset Simulation estimator $\hat{P}_f$ with respect to a given number of limit state evaluations E_T^* (compare Section 2.1), i.e.

$$\min_{p_0}\{\mathrm{CV}(\hat{P}_f) : \text{ for } E_T^* \text{ limit state evaluations}\} .$$

In general, the coefficient of variation $\mathrm{CV}(\hat{P}_f)$ is derived as (Equation 3.1 with $r = 2$ (Au and Beck (2001b)),Zuev et al. (2012))

$$\mathrm{CV}(\hat{P}_f) \approx \sqrt{\frac{1-p_0}{p_0(\log(p_0))^2}\frac{(\log p_f)^2}{E_T}(1+\bar{\gamma})} . \tag{4.1}$$

For a detailed derivation and explanation of the formula, see Au and Beck (2001b) and Zuev et al. (2012). In the following, we thus analyze (4.1) in order to optimize p_0 for implementations (I1)-(I4).

To do so, we first need to account for the underlying total number of limit state calls. If seeds are not re-used such as in methods (I1) and (I3), then we simply have $E_T = N_T = mN$. On the other hand, if we re-use seeds as samples in intermediate subsets, at subset levels $i = 1, ..., m-1$ we will already have $p_0 N$ samples of the previous subset falling into the present analyzed subset $\mathbb{F}_i$ by definition. To start the algorithm by classical Monte Carlo, we require $E = N$ limit state evaluations. Thus, when re-using the seeds, only $(1-p_0)N$ samples are left to create and evaluate in all subsets except for the first one. Again, this is due to the intersections of successive subsets including $p_0 N$ samples. In total, this results in

$$E_T = N + (m-1)(1-p_0)N = m(1-p_0)N + p_0 N$$

limit state calls when re-using seeds. For small failure probabilities (also compare (S4), Section 2.1.2) we have $m \gg 1$, so we approximately have

$$E_T \approx m(1-p_0)N_T \ .$$

We use this approximation for easier derivation and comprehensibility of the results. However, it is straightforward to substitute the approximation with the exact number of evaluations in the analysis. Note, we neglected that rejection of candidates in MCMC may require fewer evaluations if the acceptance rate of pre-candidates is less than one, e.g. in the (modified) Metropolis algorithm. This, however, will not impact our analysis at all, since we do not differ between different efficiencies of MCMC methods.

Remark 4.1.1. *Concluding, with respect to the methods in Table 4.1 we have the following total number of limit state calls:*

(I1) $E_T = mN = N_T$

(I2) $E_T \approx m(1-p_0)N = N_T(1-p_0)$

(I3) $E_T = mN = N_T$

(I4) $E_T \approx m(1-p_0)N = N_T(1-p_0)$

So, a first step for analysis of the computational demand, given by Equation 4.1 with respect to p_0, is to replace N_T by the corresponding number of evaluations E_T. This covers the discrepancy in implementation between re-using seeds for evaluation and not using samples repetitively for evaluation of different subsets. Next, we analyze $\bar{\gamma}(p_0)$ in (4.1) which is found to depend on the intermediate probabilities p_0 and has not yet been extensively analyzed under consideration of all beforehand introduced implementations (I1)-(I4).

4.1.1 Dependency of Samples Within Subsets

The dependency of samples within subsets, represented by γ_i (see Assumption 3.1.16) for subset $\mathbb{F}_i$ or the corresponding resulting average dependency of all subsets $\bar{\gamma}$, respectively, has a high impact on the coefficient of variation of Subset Simulation. It determines the effective number of independent samples, generated by MCMC in each subset. A formula for estimation after termination of Subset Simulation exists (see Assumption 3.1.16), but unfortunately its theoretical analysis is a challenging task and it also depends on the implementation of choice. Adaptive implementations aim at optimizing the MCMC sampling so that sample dependencies and therefore $\bar{\gamma}$ is minimized. On the other hand, traditional implementations do not react to the changed conditions in higher level subsets (also compare Remark 3.1.13).

Remark 4.1.2 (Higher Levels Produce More Dependent Samples)**.** *In traditional methods, a constant spread of the proposal distribution (Remark 3.1.11) is chosen. Often, the proposal spread is started at a high level which was found to yield better results than starting too low (Zuev et al. (2012)). The acceptance rate of MCMC is then typically good in the first subsets and decreases with each subset, resulting in many rejections of samples and thus a higher dependency for increasing subset levels. Evidence is given in Table 2 in Zuev et al. (2012), where the acceptance rates are found to decrease significantly with higher order subsets, so that correlations become high in high subset levels such as* $\mathbb{F}_5$ *(under* $p_0 = 0.1$*). Thus, it is plausible to assume dependencies of samples within subsets to be increasing with respect to the subset number*

$$0 = \gamma_0 < \gamma_1 < ... < \gamma_m \ .$$

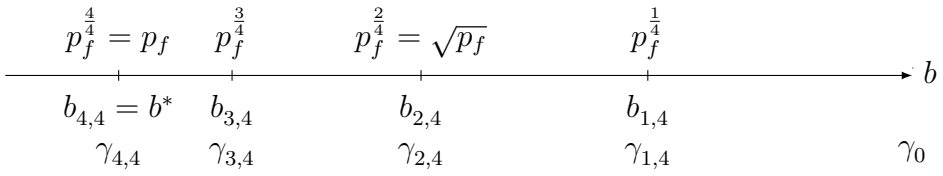
Note, within the starting set we naturally have independence, i.e. $\gamma_0 = 0$*. This is due to creation by standard Monte Carlo sampling.*

In the following, we will show how the average dependency $\bar{\gamma}(p_0)$ in traditional methods (implementations (I1) and (I2)) depends on the choice of p_0 as a consequence of Remark 4.1.2. For introduction of notation and for a better intuition on the main result, we start with an illustration of the issue and also thereby conclude, as a first finding, that $\bar{\gamma}(p_0)$ is increasing with respect to p_0. First, as we need to compare simulation runs for different choices of p_0, we need to make clear in the notation to which p_0 parameters belong.

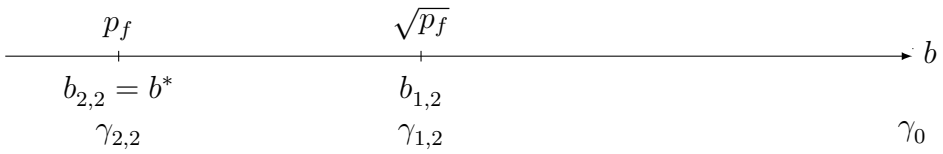
Definition 4.1.3. *For Subset Simulation with p_0 and $m = \frac{\log p_f}{\log p_0}$ integer, we define $b_{i,m}$ the thresholds of the corresponding subsets $\mathbb{F}_{i,m} := \mathbb{F}_i = \{\mathbf{x} \in D | g(\mathbf{x}) \leq b_{i,m}\}, i = 1, ..., m$ and $\gamma_{i,m}$, $i = 1, ..., m$ the correlation factors. For completeness, we also define the set $\mathbb{F}_0 := \{x \in D | g(\mathbf{x}) < \infty\}$, where $b_0 = \infty$ and $\gamma_{0,m} = 0$ by standard Monte Carlo.*

Example 4.1.4. *Given p_f, consider the following choices of p_0 in Subset Simulation:*

- $p_0 = \sqrt[4]{p_f} = p_f^{\frac{1}{4}}$:



- $p_0 = \sqrt{p_f}$:



- $p_0 = p_f$:

Observe that the p_0's are chosen so that the corresponding thresholds $b_{2,4} = b_{1,2}$ and $b_{4,4} = b_{2,2} = b_{1,1}$ coincide. This way, we may directly compare the sampling efficiency of all three approaches.

Remark 4.1.5 (Dependencies and Threshold)**.** *In general, the value of $\gamma_i, i = 0, 1, ..., m$ yields the dependency of MCMC samples in subset $\mathbb{F}_i$. However, basically, the difference between subsets is only given by their unequal threshold values b_i, so that γ_i shall directly depend on the threshold b_i and only indirectly on the subset level. It is also plausible to assume this effect to be independent of the sampling history as the MCMC itself remains mostly untouched and the seeds are distributed according to the stationary distribution directly (Remark 3.1.15). As a notation, we thus also write $\gamma(b)$ for any $b \in \mathbb{R}$.*

With Remark 4.1.5, we further conclude for Example 4.1.4 that $\gamma_{2,4} = \gamma_{1,2}$ by $b_{2,4} = b_{1,2}$ as well as $\gamma_{4,4} = \gamma_{2,2} = \gamma_{1,1}$ by $b_{4,4} = b_{2,2} = b_{1,1}$. This way, we can relate to the cases $m = 1, 2$ also by thresholds and dependencies given under $m = 4$.

Next, we look at the number of samples generated under each dependency factor. Subsets are always defined based on the p_0N samples which fall into the next subset and are used as seeds for the next subset. As a consequence, in subset $\mathbb{F}_{i+1}, i = 1, ..., m-1$, we have p_0N samples from the previous subset $\mathbb{F}_i$. These seeds can be used for evaluation of this subset too, depending on the exact implementation used (see (I1)-(I4))). The seed samples of subset $\mathbb{F}_i$ have been sampled under dependencies γ_{i-1} given in the previous subset. The missing $(1 - p_0)N$ samples are sampled with its own dependency γ_i if seeds are re-used. Otherwise, if seeds are not re-used, all N samples of subset $\mathbb{F}_i$ are sampled with dependency γ_i. Note, although only creating $(1-p_0)N$ new samples under re-use of seeds, p_0N of those will again be used for evaluation of the next subset so that we again have a total of N samples evaluated with the correlation of this subset. Thus the difference is only existent at the boundaries, which is also demonstrated in Example 4.1.6.

Example 4.1.6. *Using Remark 4.1.2, we show monotonicity of $\bar{\gamma}(p_0)$ with respect to p_0 in Example 4.1.4, also considering the different implementation cases:*

- *Seeds not used for sampling in next subset (I1):*
  - $p_0 = p_f : \bar{\gamma}(p_f) = \gamma_0$
  - $p_0 = \sqrt{p_f} : \bar{\gamma}(\sqrt{p_f}) = \frac{1}{2}\gamma_0 + \frac{1}{2}\gamma_2 > \gamma_0$
  - $p_0 = \sqrt[4]{p_f} : \bar{\gamma}(\sqrt[4]{p_f}) = \frac{1}{4}\gamma_0 + \frac{1}{4}\gamma_1 + \frac{1}{4}\gamma_2 + \frac{1}{4}\gamma_3$
- *Seeds are used for sampling in next subset (I2):*
  - $p_0 = p_f : \bar{\gamma}(p_f) = \gamma_0$
  - $p_0 = \sqrt{p_f} : \bar{\gamma}(\sqrt{p_f}) = \frac{1}{2}(1+p_0)\gamma_0 + \frac{1}{2}(1-p_0)\gamma_2 > \gamma_0$
  - $p_0 = \sqrt[4]{p_f} : \bar{\gamma}(\sqrt[4]{p_f}) = \frac{1}{4}(1+p_0)\gamma_0 + \frac{1}{4}\gamma_1 + \frac{1}{4}\gamma_2 + \frac{1}{4}(1-p_0)\gamma_3$

The differences between (I1) and (I2) for $p_0 = p_f$ and $p_0 = \sqrt{p_f}$ are trivial examples for the dependency of $\bar{\gamma}$ on the choice of p_0, regardless of seeds being re-used or not.

Although the effect shown in Example 4.1.6 might also hold for modern adaptive MCMC sampling, it is exceptionally higher for traditional methods. Comparing $p_0 = \sqrt{p_f}$ and $p_0 = \sqrt[4]{p_f}$ also shows that, for modern methods, the effect becomes negligible for higher p_0 whereas traditional methods are potentially highly affected due to the strict monotonicity and a particular steep ascent with respect to subset sample dependencies for higher order subsets. As the resulting effect directly depends on the rate of increase in $\gamma_0 < \gamma_1 < ... < \gamma_m$, we next introduce another notation to capture this feature better and also derive a realistic assumption that simplifies analysis drastically.

Definition 4.1.7 (Rate of Dependency Increase)**.** *The rate of increase of the within subset sample dependencies is given by*

$$c_\gamma(i) := \frac{1+\gamma_i}{1+\gamma_{i-1}} \ .$$

Additionally, we let

$$c_\gamma^* := \prod_{i=1}^{m} c_\gamma(i) = \frac{1+\gamma_m}{1+\gamma_0} = 1+\gamma_m$$

the maximal dependency factor of the subsets, assumed at lowest threshold $b_m = b^*$.

Remark 4.1.8 (Constant Rate of Dependency Increase)**.** *It is reasonable to consider the rate of dependency increase per subset in Subset Simulation with* $m = \frac{\log p_f}{\log p_0}$ *subsets as a constant*

$$c_{\gamma,m} := c_\gamma(1) = c_\gamma(2) = ... = c_\gamma(m) \ .$$

A constant rate is expected, since the probability of successive subsets is also set at a constant changing rate $P(\mathbf{X} \in \mathbb{F}_i) = p_0 P(\mathbf{X} \in \mathbb{F}_{i-1})$, $i = 1, ..., m$ *by definition. Thus the general structure by changing* b *in every subset, which induces* γ *directly, belongs always to the same rate of change in probability. As a consequence it is plausible that this is also reflected in the rate of dependency increase* $c_\gamma(i)$ *by MCMC sampling under* b_i, $i = 1, ..., m-1$.

Remark 4.1.8 particularly suits cases where p_f is low and the number of subsets is high, since it is easy to see that it is not correct for the switch from direct Monte Carlo in the first subset to MCMC in the second one. For higher numbers of subsets, this effect shall become negligible, but for smaller total numbers of subsets m, it is vividly present. Furthermore, by definition of c_γ^*, Remark 4.1.8 implies

$$c_{\gamma,m} = (c_\gamma^*)^{\frac{1}{m}} \ .$$

We are now able to derive a reasonable rule for assigning average dependencies to a given p_0.

Proposition 4.1.9 (Dependency of Samples Within Subsets)**.** *The factor for conversion to an effective independent sample number for evaluation under non-adaptive MCMC and no re-use of samples (I1), under assumption of integer number of subsets* m *(Assumption 3.1.5), is given by*

$$1 + \bar{\gamma}(p_0) = \frac{\log(p_0)}{\log(p_f)} \frac{c_\gamma^* - 1}{(c_\gamma^*)^{\frac{\log(p_0)}{\log(p_f)}} - 1} \ .$$

Proof. If seeds are not re-used, all N samples used for evaluation of any subset are newly created by MCMC sampling in the corresponding subset and therefore have the average dependencies of samples within this subset. Thus,

for the total number of subsets $m \geq 1$, by Definition 4.1.7 and Remark 4.1.8, we have under no re-using of seeds

$$1 + \bar{\gamma}(p_0) = \frac{1}{m} \sum_{i=0}^{m-1} (1 + \gamma_i) = \frac{1}{m} \sum_{i=0}^{m-1} \prod_{j=1}^{i} c_\gamma(j) = \frac{1}{m} \sum_{i=0}^{m-1} c_{\gamma,m}^i \ ,$$

with $c_{\gamma,m}$ as in Remark 4.1.8. This is a geometric series, resulting in

$$1 + \bar{\gamma}(p_0) = \frac{1}{m} \frac{c_{\gamma,m}^m - 1}{c_{\gamma,m} - 1} \ .$$

Using $c_{\gamma,m} = (c_\gamma^*)^{\frac{1}{m}}$, we get on the right-hand side

$$\frac{1}{m} \frac{c_\gamma^* - 1}{(c_\gamma^*)^{\frac{1}{m}} - 1} \ .$$

Then plugging in $m = \frac{\log(p_f)}{\log(p_0)}$ (Definition 3.1.2 and Assumption 3.1.5) yields the result

$$1 + \bar{\gamma}(p_0) = \frac{\log(p_0)}{\log(p_f)} \frac{c_\gamma^* - 1}{(c_\gamma^*)^{\frac{\log(p_0)}{\log(p_f)}} - 1} \ .$$

□

Under large m (resp. small failure probabilities), Proposition 4.1.9 also approximately holds for traditional implementations under re-use of seeds (I2), because, as demonstrated in Example 4.1.6, only the first and last subset are affected.

Remark 4.1.10. *In comparison to Proposition 4.1.9, modern implementations (I3,I4) use adaptive methods, thereby decreasing the increase in dependency of MCMC samples per subset. For further details, see Remark 3.1.13. Thus, changing p_0 should not significantly affect the overall $\bar{\gamma}$ in (I3,I4) with regards to Proposition 4.1.9*

$$\bar{\gamma}(p_0) \approx \bar{\gamma} = const.$$

As a matter of fact, there is a change in c_γ^* for different b^*. However, most of the time it is magnitudes lower than for traditional methods. Nevertheless, we have to remember that this is an approximation. In particular, boundary

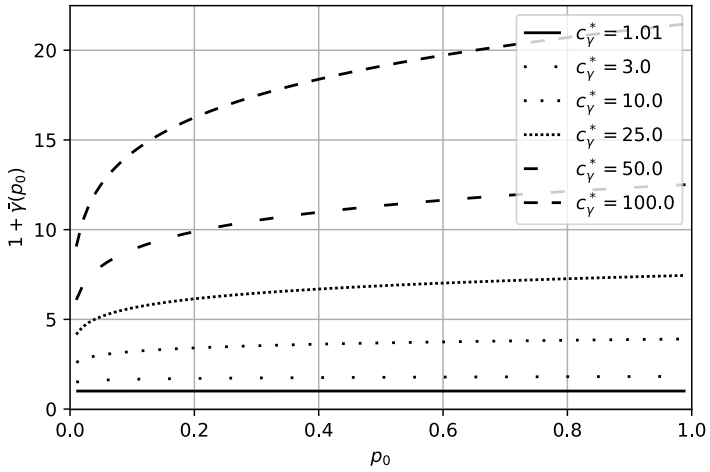


Figure 4.1: Effect of p_0 on $\bar{\gamma}(p_0)$ under different maximal MCMC dependencies c_γ^*. The failure probability was fixed to $p_f = 10^{-6}$, so that c_γ^* also refers to the MCMC sampling dependency in subset $\mathbb{F}_m = \{x \in D | g(x) < 0\}$ with $P(\mathbb{F}_m) = 10^{-6}$. Keep in mind that an MCMC sampling corresponding to this maximum dependency c_γ^* is actually not performed in Subset Simulation, because when reaching the last subset no additional MCMC sampling is required.

cases such as $p_0 = p_f$, having $\bar{\gamma} = 0$ (standard Monte Carlo), will always result in a bad approximation as we are generally far from perfect sampling in MCMC.

Concluding, the factor for conversion to an effective independent sample number for evaluation under traditional and modern MCMC sampling is given by

$$\text{(I1,I2):}\ 1 + \bar{\gamma}(p_0) = \frac{\log(p_0)}{\log(p_f)} \frac{c_\gamma^* - 1}{(c_\gamma^*)^{\frac{\log(p_0)}{\log(p_f)}} - 1}$$

$$\text{(I3,I4):}\ 1 + \bar{\gamma}(p_0) \approx 1 + \bar{\gamma} = \text{const.}$$

An illustration of the result is given in Figure 4.1, not considering the trivial cases (I3) and (I4). The effect seems not to be very strong, but still allows a better comparison of traditional and modern methods as well as an improved understanding of the mechanisms of the Subset Simulation algorithm. Additionally, if p_f becomes low, we typically have very high c_γ^* in non-adaptive

MCMC sampling so that it must definitely be considered then. For example, if $c_\gamma^* = 25$ then $\frac{1+\bar{\gamma}(0.1)}{1+\bar{\gamma}(0.6)} = 0.80$ and for $c_\gamma^* = 100$ even $\frac{1+\bar{\gamma}(0.1)}{1+\bar{\gamma}(0.6)} = 0.73$. Simulations haven shown that such or even higher values are not unrealistic in traditional MCMC sampling (see e.g. Zuev et al. (2012)), and appear frequently for low enough failure probabilities due to the fixed proposal spread which then leads to rejection of many samples in MCMC sampling. Also note that c_γ^* does not belong to a subset that is really evaluated in Subset Simulation. Indeed, it belongs to b^*, the failure set where the algorithm stops and no MCMC is performed. The quotient $\sqrt{\frac{1+\bar{\gamma}(0.1)}{1+\bar{\gamma}(0.6)}}$ then directly affects the relation of the coefficients of variation. Continuing above exemplary relations, we have an effect on coefficients of variation of $\sqrt{0.8} = 0.89$ and $\sqrt{0.73} = 0.85$. Equivalently we could also state that we require 0.8 and respectively 0.73 as much limit state evaluations as assumed before to achieve the same accuracy in estimation since the change in average dependencies $\bar{\gamma}(p_0)$. This is in favor of low p_0 and yields one explanation for discrepancies between statistical analysis and simulation (compare Section 3.2), particularly for traditional SuS implementations (I1,I2).

4.1.2 Optimal Choice: Result

We put together our results in the following Theorem.

Theorem 4.1.11 (Coefficient of Variation: Based on Implementation)**.** *The coefficient of variation is approximately given by*

$$\mathrm{CV}(\hat{P}_f, E_T) \approx \sqrt{\frac{(1-p_0)^r}{p_0(\log(p_0))^2}\frac{(\log p_f)^2}{E_T}(1+\bar{\gamma}(p_0))}$$

with

$$1+\bar{\gamma}(p_0) = \frac{\log(p_0)}{\log(p_f)}\frac{c_\gamma^* - 1}{(c_\gamma^*)^{\frac{\log(p_0)}{\log(p_f)}} - 1} \ .$$

for traditional implementations (I1,I2) and $\bar{\gamma}(p_0) \approx \bar{\gamma} \in [0,\infty)$ for modern methods (I3,I4). Parameter r is set to $r = 1$ if sample seeds are not-reused (I1,I3) and $r = 2$ if they are re-used (I2,I4).

Proof. Follows directly by plugging the number of limit state evaluations (Remark 4.1.1) and the dependency factor with respect to p_0 (Proposition 4.1.9, Remark 4.1.10) into Equation 4.1. □

Minimizing $\mathrm{CV}(\hat{P}_f, E_T)$ with respect to p_0 is thus achieved by minimizing the corresponding approximation of $\mathrm{CV}(\hat{P}_f, E_T)$ for the chosen implementation according to Theorem 4.1.11. Results are then easily derived by Figure 4.2. It is easy to see that the currently used and considered as best parameter selections in the interval $[0, 1, 0.3]$ (cf. Zuev et al. (2012)) are robust to the chosen implementation, but on the other hand are not always the best ones according to Theorem 4.1.11, in particular for the most efficient method (I4). This raises the question why high p_0 are typically not used yet, even in modern implementations. The reason is twofold. First, low p_0 values were found to outperform high p_0 in the past where only methods (I1-I2) were present. Second, simulations did not yet manage to reproduce the low coefficients of variation for $p_0 > 0.3$ as given in the illustration (compare e.g. Au and Wang (2014), and Li and Cao (2016) even identifies $p_0 = 0.3$ as problematic). Due to the high popularity of the algorithm, recommendations to use low p_0 were justified in many simulations, yielding undisputed evidence for correctness, at least if sticking to the implementations used in the past. Still, we need to find an answer on why theoretical findings could not yet be validated in past simulations. The reason has to be the consequence of a counteracting effect that is not considered in the analysis, which again raises the question whether it can be prevented or not. Indeed, we will identify such an effect and show that the former holds true. First, however, we want to remark that Theorem 4.1.11 is based on some assumptions possibly not fulfilled, as stated in the following Remark and Section 4.1.3.

Remark 4.1.12 (Modern Implementations: Constant $\bar{\gamma}$?)**.** *The average dependency of samples within subsets $\bar{\gamma}$ is assumed independent of p_0 under adaptive MCMC. However, it is likely that also for modern methods the dependency structure as described for the traditional methods still applies in a weak form.*

Indeed we also find evidence for partially falsifying the assumption of a constant $\bar{\gamma}$ in Section 4.1.3 with respect to p_0 by simulation, even beyond the facts mentioned beforehand. Nevertheless, other findings are found even more important and underline the general result of a flat coefficient of variation curve above $p_0 \geq 0.1$ (also above 0.3), so that also higher p_0 values generally lead to good efficiency. In the next section, we identify the naturally altered chain length in current implementations of SuS as the main reason for the high discrepancies between theory and observation in simulations. This also explains the high variations of the SuS results for high

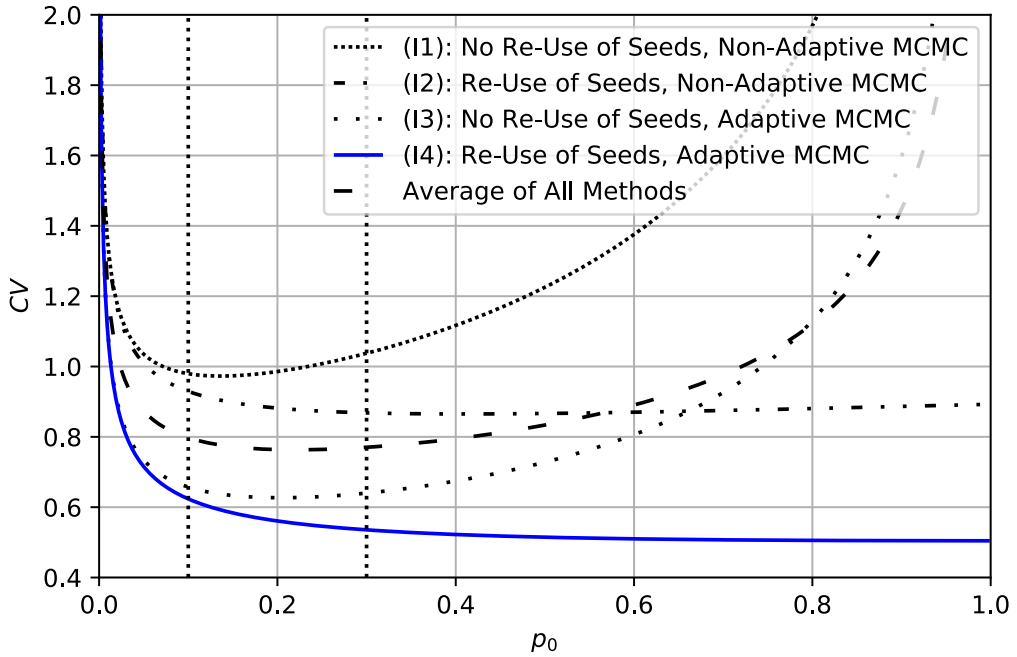


Figure 4.2: Comparing the Coefficient of Variation (CV) for all four implementations (I1-I4). The failure probability is fixed to $p_f = 10^{-6}$ and dependency capturing parameters are set to $c_\gamma^* = 50$ in (I1,I2) and $\bar{\gamma} = 3$ in (I3,I4). We additionally marked the interval $[0.1, 0.3]$ which is currently considered to be the best choice for p_0. The currently most efficient implementation available (I4) is highlighted by blue color and solid line. Building the average value over all implementations aims at demonstration of an always acceptable choice of p_0 when several implementations are considered to be in use.

p_0, because the chain length (as given in Definition 3.1.3) directly depends on it in current implementations. The relation is given by $N_l = \frac{N}{N_c}$ with $N_c = p_0 N$ so that $N_l = \frac{1}{p_0}$, where for $p_0 > 0.5$ and under re-use of seeds we may not use all the potential seeds anymore so that we are required to choose some of them at random for sample production. However, the same procedure can be applied for lower $p_0 \leq 0.5$, then requiring to artificially increase the chain lengths to receive N samples again. In Remark 4.1.13, we define the resulting approach without loss of generality only under re-use of seeds. Conventionally, seeds are re-used for an increased efficiency of the algorithm.

Remark 4.1.13 (Artificially Increasing the Chain Length)**.** *[Compare Au and Wang (2014), page 174 "More or Fewer Chains"] We allow to pick and use only a fraction* $c_s \in (0, 1]$ *of the potential number of seeds* $\min\{p_0 N, (1 - p_0)N\}$ *as initial chain state, resulting in* $N_c = c_s p_0 N$ *chains for* $p_0 \in (0, 0.5]$ *and* $N_c = c_s(1 - p_0)N$ *for* $p_0 \in (0.5, 1)$ *. The corresponding chain length is* $N_l = 1 + \frac{(1-p_0)N}{N_c}$ *(also counting the seed sample),*

$$N_l = 1 + \frac{1}{c_s} \frac{1 - p_0}{p_0}$$

for $p_0 \in (0, 0.5]$ *and*

$$N_l = 1 + \frac{1}{c_s}$$

for $p_0 \in (0.5, 1)$*.*

The opportunity to generate more or fewer chains in this manner has previously been mentioned in Au and Wang (2014), but to the author's best knowledge never been investigated further. The idea why this approach might improve efficiency considerably, is that for example $p_0 = 0.5$ results in a natural chain length of $N_l = 2$, thus only one new generated sample for every seed. If we think of a tree structure with the seeds in the first subset identified as roots, the distance of seeds will then be very low in comparison to the case when longer chains are used. This low average distance again implies a high autocorrelation by MCMC and thus results in high correlations of the samples. In higher level subsets, this might increase correlations significantly, especially when double appearances of samples become likely. Also the independence assumption between chains, often made for estimation of correlations, is rather admissible for longer chains.

4.1.3 Simulation Study

We want to add steps towards consensus of theory and simulation. Also, as our aim is to explore the most efficient method and its optimal parameters, we restrict ourselves to the most efficient implementation ($I4$), which was found to outperform the other approaches (see Papaioannou et al. (2015)). Our findings will also have the greatest effect on (I4). Reusing seeds is clearly favorable as they provide information at no cost and there is consent in doing so. In the following simulation study and intermediate discussion, we aim at verification or falsification of the theoretical findings by practical simulations. Although we can of course not give certain results due to the exemplary character of the given simulations, its results will give evidence and hold approximately in many situations. This will help us to possibly identify wrong assumptions and to further optimize the algorithm by proper parameter choices matching theoretical and practical findings. We restrict ourselves to $p_0 = 0.1, 0.3, 0.5, 0.8$, because this will most likely cover the most relevant cases sufficiently while keeping computational efforts in simulation low.

For demonstration, we consider examples 'Example 1' (linear), 'Example 2a' (convex) and 'Example 2b' (concave) given in Papaioannou et al. (2015) (also compare Fujita and Rackwitz (1988)). The corresponding limit state functions are given by

$$g_1(\mathbf{X}) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{X}_i + \beta \tag{4.2}$$

$$g_{2a}(\mathbf{X}) = \eta_a - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{X}_i \tag{4.3}$$

$$g_{2b}(\mathbf{X}) = -\eta_b + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{X}_i \tag{4.4}$$

where in g_1, $\mathbf{X}$ is a vector of independent standard normally distributed random variables and in g_{2a} and g_{2b}, it is a vector of independent exponentially distributed variables. The constants β, η_a and η_b allow to control the failure probability. Because we know that the first subset is produced by independent samples, a high weighted first subset might distort the results. Thus we choose $p_f = 1 \cdot 10^{-8}$, a very low failure probability, for simulation. Further information is given in the discussion in Section 4.1.4. Table 4.2 shows the

results generated by implementation (I4), using the classical chain length of $N_l = \frac{1}{p_0}$. Results by simulation in Table 4.2 contradict our theoretical statement that for the adaptive methods p_0 becomes flat from 0.3 on and is generally not varying much with respect to p_0. The bias appears to be low, although not necessarily negligible, and similar for all p_0. Now, although this appears to reject our theoretical statements, we also see that the chain length is decreased when p_0 is increased. So indeed we also simulated the effect of an altered chain length. This was a natural consequence of $N_l = 1 + \frac{(1-p_0)}{p_0}$, following by the demand of constant N in every subset when re-using and utilizing all seeds for MCMC in intermediate subsets $N_c = p_0 N$. The idea now is that for similar chain lengths, the dependencies of samples generated by MCMC might be similar, regardless of p_0. Simulations of Figure 4.2 are extended in Table 4.3 to examine this hypothesis.

Table 4.3 shows that the bias is increased for longer Markov chains, generally resulting in a higher coefficient of variation unless the bias is corrected afterwards. The reason is unclear at the moment, but Section 4.2 will bring some clarity (The bias created as described in and around Remark 4.2.8 could be the reason). Nevertheless, simulations give evidence that the efficiency of Subset Simulation is insensitive with respect to the choice of $p_0 \in [0.1, 0.8]$ if the chain length is adapted properly. When adapting, it however is important to remember that increasing the chain length does reduce the number of chains so that ergodicity issues might become more relevant (cf. Au and Beck (2001b), Au and Wang (2014)). Care has to be taken. Thus, we will next also look at a higher number of samples to avoid the upcoming bias and compare the results for more different choices of the chain length while restricting ourselves to the linear limit state case as differences between the limit state types appeared to be insignificant. Our aim is to figure out at which point further increasing of the chain length will not lead to significant efficiency improvements anymore. The results are unbiased up to 1.5% in the worst case and generally smaller than 1% so that we refrain from showing the estimated failure probabilities. Table 4.4 clearly shows that the choice of the chain length can be optimized and has a significant, for high p_0 one of the most important ones, effect on efficiency. Also interesting is that for high p_0 the coefficient of variation was found to slightly increase again for very high chain lengths. An illustration of the results in Table 4.4 is also given in Figure 4.3. Under the premise that a higher number of chains is favorable due to wanted ergodicity of the algorithm, one has to find a good balance

	$\mathbf{p_0}$	**0.1**	**0.3**	**0.5**
	N	1000	650	520
	p_a	0.1	0.2	0.4
	N_l	10	≈ 3	2
$\hat{\mathbf{p}}_\mathbf{f} \cdot 10^{-8}$	linear ls	1.08	1.07	1.05
	convex ls	1.05	1.06	1.03
	concave ls	1.11	1.06	1.05
CV	linear ls	0.58	0.81	1.21
	convex ls	0.62	0.75	1.07
	concave	0.64	0.84	1.06
CV*	linear ls	0.54	0.74	1.13
	convex ls	0.58	0.69	1.02
	concave	0.58	0.77	0.99

Table 4.2: **Efficiency comparison under varied $\mathbf{p_0}$ and chain length $\mathbf{N_l}$ for different limit state types.** Results are based on $10,000$ independent simulation runs for linear and 2000 runs for convex and concave limit state functions with implementation (I4). The total number of evaluations E_T under re-use of samples is approximately kept constant for all the choices, resulting in different N with respect to p_0. The adjusted coefficient of variation CV* is calculated with respect to the empirical means instead of the true failure probability $p_f = 1 \cdot 10^{-8}$ and is also corrected to an equal total number of samples. This is an interesting comparison because the bias might possibly be removed by proper sampling or transformation after termination of the algorithm (cf. Cérou et al. (2012) and Section 4.2). At the same time we try to keep the number of total proposal spread updates similar for all the methods (c.f. p_a, which gives the percentage of chains after which adaption is performed). Underlying limit state function and basic variable distributions are as in Papaioannou et al. (2015) Example 1 (linear), Example 2a (convex) and Example 2b (concave) with dimension $d = 10$. The first column belongs to the typical recommended choice of p_0 and chain length.

	$\mathbf{p_0}$	**0.1**		**0.3**		**0.5**		**0.8**	
	N	1000		650		520		410	
	p_a	0.1		0.2		0.4		0.6	
	c_s	1	0.1	1	0.1	1	0.05	0.25	0.01
	N_l	10	91	3	24	2	21	2	26
$\hat{\mathbf{p}}_\mathbf{f} \cdot 10^{-8}$	linear ls	1.08	1.12	1.07	1.11	1.05	1.12	1.04	1.14
	convex ls	1.05	1.12	1.06	1.09	1.03	1.12	1.06	1.16
	concave ls	1.11	1.11	1.06	1.11	1.05	1.10	1.10	1.13
CV	linear ls	0.58	0.55	0.81	0.56	1.21	0.58	0.94	0.60
	convex ls	0.62	0.58	0.75	0.58	1.07	0.60	1.01	0.63
	concave	0.64	0.55	0.84	0.59	1.06	0.57	1.27	0.66
CV*	linear ls	0.54	0.49	0.74	0.50	1.13	0.50	0.87	0.51
	convex ls	0.58	0.52	0.69	0.52	1.02	0.52	0.92	0.53
	concave	0.58	0.50	0.77	0.52	0.99	0.51	1.12	0.57

Table 4.3: **Efficiency comparison under varied $\mathbf{p_0}$ and chain length $\mathbf{N_l}$ for different limit state types.** This table is an extension of Table 4.2. Settings are exactly the same, where additional columns for varied chain length as well as for $p_0 = 0.8$ have been added. The chain length is an exact consequence of the choice of c_s as described in Remark 4.1.13. We see that a similar coefficient of variation is reached for all p_0 when an appropriate chain length is selected, where the coefficients of variation generally depend highly on the chain length.

$\mathbf{p_0}$	**0.1**	**0.3**	**0.5**	**0.8**
N	10,000	6500	5200	4100
p_a	0.1	0.2	0.4	0.6
c_s	CV* (N_l)			
1	0.161(10)	0.207(3)	0.271(2)	-
0.5	0.158(19)	0.170(5)	-	-
0.25	0.152(37)	0.156(10)	0.172(5)	0.273(2)
0.1	0.153(91)	0.149(24)	0.153(11)	0.176(3)
0.01	0.151(901)	0.149(234)	0.150(101)	0.149(26)
0.001	-	-	0.160(1001)	0.159(251)

Table 4.4: **Efficiency comparison under varied $\mathbf{p_0}$ and chain length $\mathbf{N_l}$: Searching for the optimal chain length.** Results are based on 2000 independent simulation runs. The chain lengths assumed to be optimal are marked with the darker blue color. Furthermore, typical selected chain lengths are marked with bright blue color, demonstrating that for high p_0 it is very important to alter the chain length and not use the standard selection.

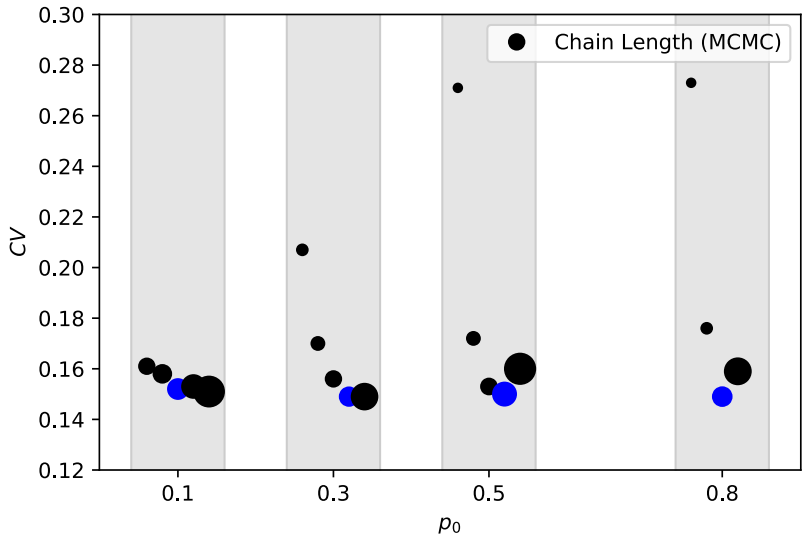


Figure 4.3: Efficiency comparison under varied p_0 and chain length: A bigger size of the markers reflects a higher chain length. Possible choices for an optimal chain length are marked blue.

between ergodicity (cf. Remark 3.1.14) and chain length. In our example, we might choose a chain length of $N_l = 37$ for $p_0 = 0.1$, $N_l = 24$ for $p_0 = 0.3$, $N_l = 11$ for $p_0 = 0.5$ and $N_l = 26$ for $p_0 = 0.8$. In general, less seed samples might be sufficient for start of MCMC concerning ergodicity (Remark 3.1.14) for higher p_0. According to Theorem 4.1.11, the convergence of the CV is of order $\frac{1}{\sqrt{E_T}}$. Thus, in the case of for example picking $p_0 = 0.5$, a change from $N_l = 2$ to $N_l = 11$ approximately yields the same accuracy for a relatively reduced to $32\% = \left(\frac{0.153}{0.271}\right)^2$ number of total evaluations E_T, related to the original number of total evaluations. This is a tremendous increase in efficiency and makes it possible to use high intermediate subset probabilities p_0 with the same efficiency as low p_0.

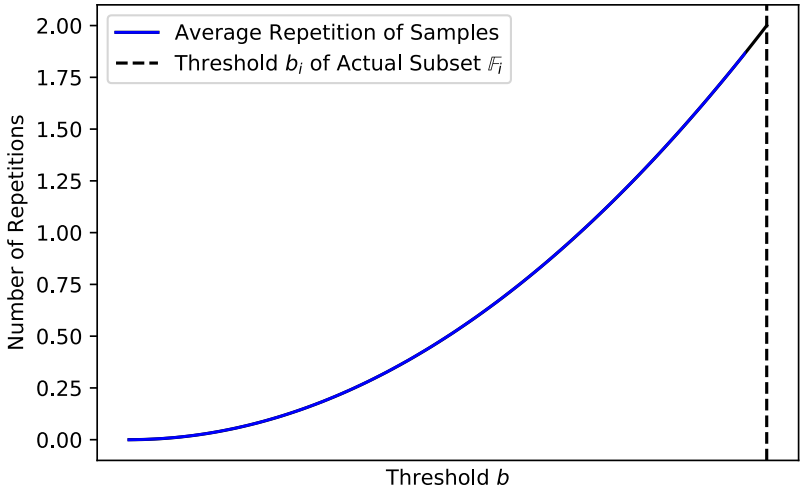
While this is a positive effect, the question remains as to why our expectations for large N were not met. Also choosing high values of p_0 still does not bring a clear advantage over lower p_0 values but only similar coefficients of variation. Here a look at the dependencies within subset samples $\bar{\gamma}(p_0)$ provides a good reason. For example picking $p_0 = 0.1$ and $p_0 = 0.5$ leads to corresponding empiric within subset sample dependencies $\bar{\gamma}(0.1) \approx 1.7 - 2.0$ and $\bar{\gamma}(0.5) \approx 2.7 - 3.0$, respectively. This statement is based on examinations with $10,000$ independent repetitive runs of Subset Simulation under varied $p_f \in \{1 \cdot 10^{-3}, 5 \cdot 10^{-4}, 5 \cdot 10^{-5}, 1 \cdot 10^{-6}, 5 \cdot 10^{-7}, 5 \cdot 10^{-8}, 1 \cdot 10^{-8}\}$ and $N \in \{500, 1000, 5000\}$, $N \in \{300, 600, 3000\}$ respectively for $p_0 = 0.1, 0.5$ as well as some variation of the limit state function (linear (g_1), convex g_{2a}, concave g_{2b}). To understand this effect, the construction of the subsets in Subset Simulation returns a plausible reason.

- New samples are generated by MCMC, where samples that exceed the threshold of the subset are rejected. This implies that samples which start close to the threshold are rejected more frequently and are stronger correlated in general. This should result in a dependency structure similar as the one given in Figure 4.4.

- The next threshold is chosen at the p_0N-th lowest limit state value of the samples of the subset. For high p_0, this is the limit state value of a sample close to the threshold and for low p_0 it is a sample's limit state value which is generally far from the threshold.

Now both points taken together imply that the derivation of the threshold tends to be based on less dependent samples for low p_0 than for high p_0. An illustration is given in Figure 4.4. The next remark summarizes this aspect.

Figure 4.4: Expected Doubles by MCMC Samples: An Illustration of a reasonable structure of sample repetitions by MCMC in Subset Simulation. The frequency of sample repetitions varies with respect to the threshold and approaches its maximum at the threshold of the subset.



Remark 4.1.14. *Samples used for determining the threshold for small p_0 are only slightly correlated as those are based on the 'flat' part of the available samples. As a result, $\bar{\gamma}(p_0)$ could depend heavily on p_0, even beyond the strong impact of the chain length selection. This seems to be in contrast to Theorem 4.1.11 and state of the art (Section 3.1), but in accordance with simulations. Thus bias and variance increase for higher p_0 with respect to the derived formulas derived here and in the past are not perfect for optimization of p_0 without the extension to a variable $\bar{\gamma}(p_0)$. However, even if this variation is included in $\bar{\gamma}(p_0)$, the formulas do not lead to correct results. We think that this is due to the rather problematic analysis of order statistics with dependencies, as discussed in the next section.*

The suspected behavior in Figure 4.4 was also validated by simulations, where typical results are shown in Figure 4.5 and Figure 4.6. Different choices of p_0 did not significantly change the structure of threshold frequencies in our experiments. Furthermore, the shape of the limit state function was found to possibly have some impact on the result, but most likely only a minor one. This makes it hard to draw general conclusions on the behavior of $\bar{\gamma}(p_0)$ with respect to p_0. In particular it is also not clear if empirically

Figure 4.5: Frequency of subset samples with equal limit state values, indicating the corresponding correlations with regards to selection of the subset threshold. The result is based on a Subset Simulation run with $p_0 = 0.1$ and $N = 1000$ with respect to a linear limit state function. Four successive subsets are considered. For comparison, we also marked where $p_0 = 0.5$ would have set the threshold of the next subset, to show that $p_0 = 0.1$ selects the next subset threshold from a part of the domain which has comparably few doubles, yielding evidence for less correlated samples and therefore a comparably higher number of effective samples. This is a feasible approach because the subset sample structure looks similar under $p_0 = 0.5$. The frequency of a threshold b, if selected, coincides with the number of repetitive limit state values of value b appearing in the current subset.

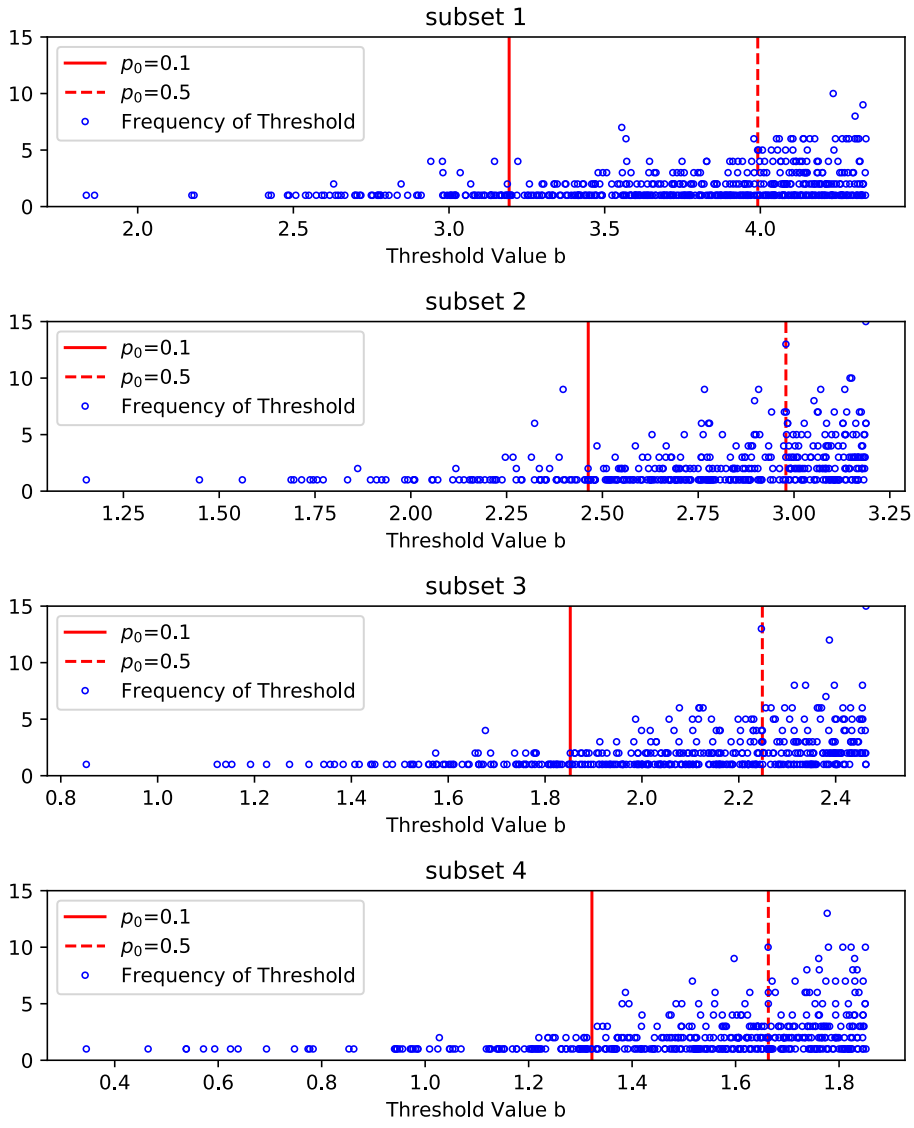
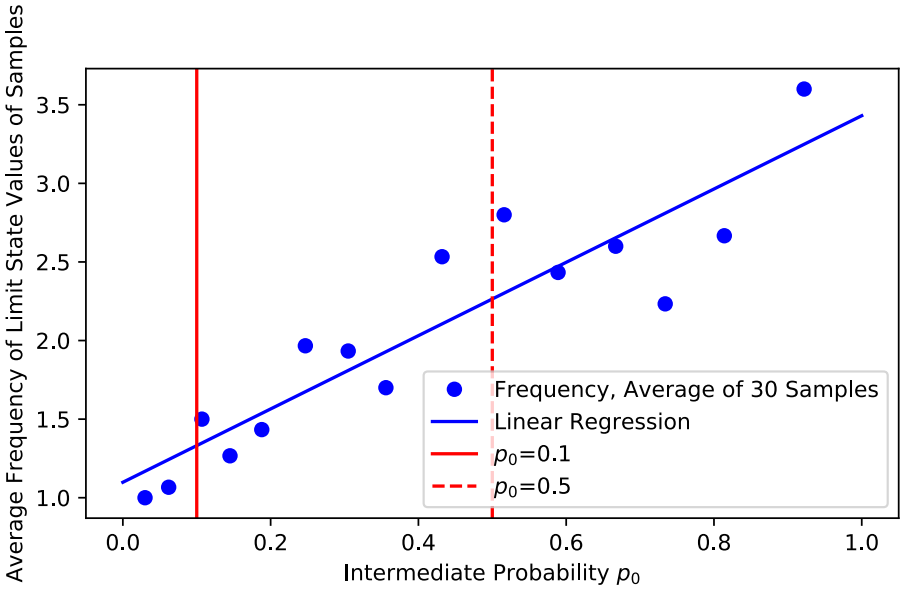
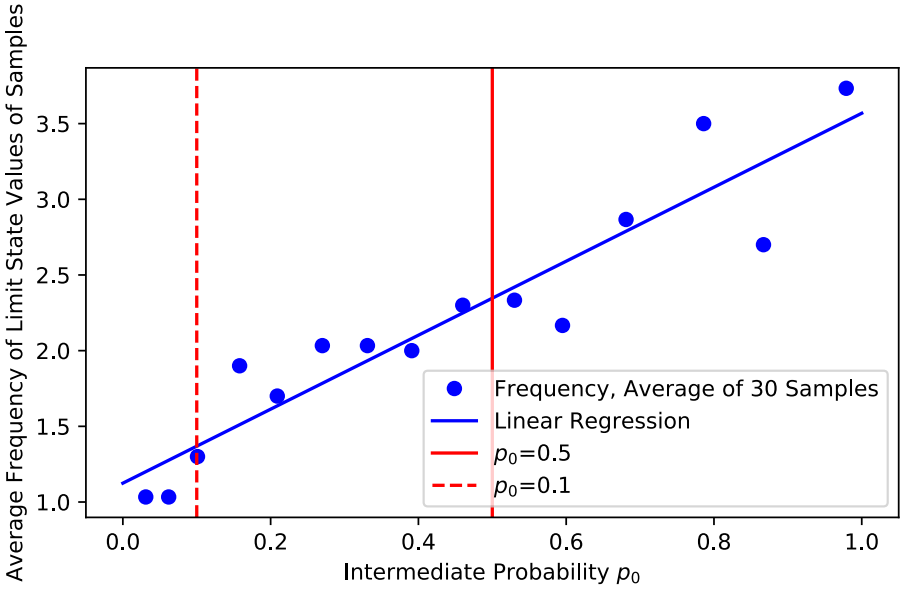


Figure 4.6: Average frequency of limit state values, or equivalently samples, with respect to the intermediate probability p_0. Both examples are for a linear limit state function (Equation 4.2, g_1). For convex and concave limit state functions, no significant differences were observed.



(a) Samples generated with $p_0 = 0.1$, $N = 1000$.



(b) Samples generated with $p_0 = 0.5$, $N = 600$.

derived autocorrelations allow to draw straightforward conclusions for the coefficient of variation since complexity of dependency in order statistics goes beyond that. For analysis this could be a relevant drawback, but for practical application it can be assigned a bit less importance. Appropriate values for correlations with respect to p_0 can still be approximately derived by empirically evaluating the autocorrelation of the Markov chains and yield satisfying enough results. Remark 4.1.14 explains why we might have discrepancies between different p_0 even for the same reliability evaluation. Unfortunately, this change of $\bar{\gamma}(p_0)$ with respect to p_0 compensates our conclusion of profiting from choosing higher p_0 values so that we do not increase efficiency much when choosing high p_0. Nevertheless, the general result that high p_0 can be beneficial still holds. Concluding, the results by simulation show that, under appropriate Markov chain lengths, we find insensitivity of the coefficient of variation to the choice of p_0 as stated in the statistical analysis. Yet, we still see that theoretical results and simulation outcomes do not coincide perfectly. A plausible reason for this outcome was given but a full explanation is prevented by unknown properties of dependent order statistics. Some additional causes are also discussed in the following section.

4.1.4 Discussion

We found evidence for analytically derived results on the optimal choice of p_0 in modern implementations by simulations. Still, there are discrepancies which were not yet explained. These might come from assumptions that do not stand the real world settings. Some shortcomings of the analysis in this section, also beyond Section 3.2, are discussed in the following.

Remark 4.1.15 (The Special Character of the First Subset)**.** *Our derivation of the main results is based on a constant proportion* $\frac{1+\gamma_{i+1}}{1+\gamma_i}, i = 1, ..., m-1$ *(I1,I2) or a constant* $(1+\gamma)$ *for all subsets (I3,I4). This assumption is violated clearly with regards to the first subset estimation by Monte Carlo, where* $\gamma_0 = 0$*, so that we will have a possibly non-negligible error. In fact, this error increases for higher failure probabilities and also for high* p_0*. Then the effect by using less independent samples by standard Monte Carlo in the first subset can become significant. In general this effect will always reduce the efficiency of higher* p_0 *in comparison to lower* p_0*.*

We observe a reduced coefficient of variation when choosing higher p_0 in general, but can also increase efficiency by using more standard Monte Carlo

samples, i.e. choosing a low p_0. To benefit from this, we suggest to fix the first intermediate subset probability by Monte Carlo at $p_0 = 0.1$ and then switch to another $p_0 \geq 0.1$, for example $p_0 = 0.5$, afterwards. Then we may approximate the relations of different choices of p_0 by Theorem 4.1.11 even for higher failure probabilities, without having a noticeable approximation error by the violation of the assumptions with respect to $(1 + \gamma)$ in the first subset.

Remark 4.1.16 (Thresholds)**.** *The implementation of Subset Simulation suffers from an inappropriate definition of the threshold and seeds of the subsets. For high p_0, typically smaller N are chosen to keep the total number of samples in comparison to lower p_0, so that the number of subsets m naturally increases. These two conditions together, result in a much stronger impact of this effect for high p_0.*

A stochastically appropriate threshold selection is examined in Section 4.2, also discovering a relation of it to the resulting bias in Subset Simulation, which could not be explained in simulations yet. Conclusions are summarized in Section 4.3, together with the findings of the next section.

4.2 Novel Stochastic Interpretation and Explicit Analysis

Outline

In this section, we analyze Subset Simulation beyond asymptotic results, introducing and utilizing a new stochastic interpretation of the algorithm. This allows for an increased understanding of the algorithm in general and also for an explicit bias correction. The basis of our analysis is the interpretation of real intermediate subset probabilities as an order statistic with stochastic threshold values, the local (subset-wise) perspective, and the combination of Bayes and Frequentist analysis. Based on the extended stochastic analysis, we can prove independence of successive intermediate subset probabilities under weak conditions and discuss an appropriate successive threshold selection for defining the subsets. The consequences of independence are far-reaching, resulting in a more substantial argumentation basis for Subset Simulation and a good outlook for further discoveries in the future. For an easier interpretation of the algorithm, we also change the seed selection for MCMC, resulting in a new implementation. Evidence on theoretical findings is given in simulations.

Preliminaries

Before we start with the analysis, let us briefly recall the concept of order statistics and also describe how we proceed in this section.

Remark 4.2.1 (Order Statistics)**.** *Order statistics deal with the distribution of ordered random samples. In general, given N iid samples $z_1, ..., z_N$, in ascending order, the k-th order statistic is the k-th element, $0 < k \leq N$ of the ordered samples $z_{(1)} \leq z_{(2)} \leq ... \leq z_{(N)}$, thus the k-th smallest value $z_{(k)}$. A detailed introduction to the topic is given in Arnold et al. (1992).*

Using order statistics is necessary to guarantee a suitable assignment of intermediate subset probabilities and thresholds. More information is given in Remark 4.2.31. Due to the difficulties in analysis of dependent order statistics, we have to stick to Assumption 3.1.10 (independent within subset samples). However, when it is possible to avoid it, we will do so. Otherwise dependencies are included in the analysis later, first assuming independence

and then converting the number of dependent samples to a representative number of effective independent samples. This approach is typical for Subset Simulation (cf. Au and Beck (2001b)).

Analysis of Subset Simulation is generally based on one of the two following approaches. First, one can assume a fixed probability of failure as given and examine the properties of estimations that arise from applying Subset Simulation on the given problem. This approach is good for prediction of performance and therefore can be utilized for parameter optimization. In the following, we will also refer to it as the Frequentist approach since it is indeed a Frequentist stochastic setting, having a fixed true failure probability and considering the resulting estimates as random. Examples in literature include Au and Beck (2001b) and Cérou et al. (2012). We choose the second approach, stated in Definition 4.2.2.

Definition 4.2.2 (Setting for Analysis: Bayesian Approach)**.** *The real probability of failure p_f is assumed to be unknown, considered as a stochastic variable. In contrast to p_f, its estimate $\hat{p}_f$ is given. Thereby, we are given the number of subsets $m \in \mathbb{N}$ and the corresponding intermediate probabilities $\hat{p}_i = p_0 \in (0,1), i = 1, ..., m-1$, $\hat{p}_m \in (p_0, 1]$. This is a Bayesian approach, assuming the true failure probability as stochastic.*

As explained in Zuev et al. (2012), who follow a Bayesian approach as in Definition 4.2.2, it is possible to derive a probability distribution of p_f after given a Subset Simulation result. In their work, the result is referred to as 'Bayesian Post Processor'. This is exceptionally practice oriented and allows to derive confidence intervals for the failure probability which is a typical demand in reliability evaluation under stochastic evaluation methods. Besides its practical relevance, the Bayesian approach is also chosen in favor of the Frequentist because its analysis is more straightforward, avoiding to estimate m and $\hat{p}_m$.

Similar as Zuev et al. (2012), we also follow the Bayesian approach most of the time, but in addition take a local perspective with order statistics on Subset Simulation and in particular focus on suitable results for small N. Importance of such analysis is discussed in Remark 3.2.1. Nevertheless, results also allow new conclusions such as independence between intermediate subset probabilities and a new understanding of dependencies of samples within subsets, for high sample numbers.

4.2.1 Stochastic Interpretation by Order Statistics

First, we describe the construction of Subset Simulation subsets and show how it is directly related to order statistics, giving a basis for clear reference in the following. Analysis of Subset Simulation by order statistics has already been proposed in Cérou et al. (2012), but unlike ours, results are based on asymptotic arguments and are in the Frequentist setting.

Definition 4.2.3 (Construction of Subsets)**.** *The construction of subset* $\mathbb{F}_i, i = 1, ..., m-1$ *is as follows:*

1) *Produce* N *samples* $x_1^{r,i}, ..., x_N^{r,i} \in \mathbb{F}_i$*, with corresponding limit state values* $g_1^{r,i}, ..., g_N^{r,i}$*, by Monte Carlo (*$i = 0$*,* $b_0 = \infty$*) or Markov Chain Monte Carlo (MCMC) (*$i \geq 1$*) using the* $(p_0 N - \zeta)$*-th lowest values of the previous subset* $\mathbb{F}_{i-1}$ *as seeds, naturally falling into* $\mathbb{F}_i$*. Here* $\zeta \in \{1, ..., p_0 N\}$ *is the number of the* $p_0 N$ *lowest samples equal to the threshold* b_i *(also compare Remark 4.2.7 for more details).*

2) *Sort the limit state values of the samples* $g_1^{r,i}, ..., g_N^{r,i}$ *in ascending order* $g_{(1)}^{r,i}, ..., g_{(N)}^{r,i}$*. The corresponding samples are ordered likewise.*

3) *Set the threshold* b_{i+1} *equal to the* $p_0 N$*-th lowest limit state result* $b_{i+1} := g_{(p_0 N)}^{r,i}$*.*

4) *Define the next subset with the given threshold* $\mathbb{F}_{i+1} := \{\mathbf{x} \in D | g(\mathbf{x}) \leq b_{i+1}\}$*. If* $b_{i+1} < b^*$ *set* $b_{i+1} = b^*$ *and stop; termination in subset* m *is reached.*

Subset m is an exception as it has a fixed threshold $b_m = b^*$ and switching from the standard evaluation to evaluation of the last subset is a stopping criterion, repeating $1) - 3)$ until $b_i \leq b^*$ for an $i \geq 1$. Step 1) in our construction of subsets is in contrast to conventional implementations insofar as we do not use samples with values equal to the threshold of a subset as seeds. We can also interpret the construction of subsets with regards to the Frequentist setting.

Remark 4.2.4 (Identifying Stochastic and Deterministic Values in the Frequentist Setting)**.** *An important observation in Definition 4.2.3 is that the stochastic variables in Subset Simulation are the thresholds* $b_1, ..., b_{m-1}$ *which*

define the individual subsets except for the last one which is set to $b_m = b^*$. *As a consequence, the real intermediate subset probabilities*

$$P_{r,i} = P(\mathbb{F}_i|\mathbb{F}_{i-1}) = P(g(\mathbf{X}) \leq b_i|g(\mathbf{X}) \leq b_{i-1}), i = 1, ..., m-1$$

are also stochastic. The corresponding estimated probabilities $\hat{p}_i = \hat{P}(\mathbf{X} \in \mathbb{F}_i|\mathbf{X} \in \mathbb{F}_{i-1}), i = 1, ..., m-1$ *on the other hand are deterministic and given by* $\hat{p}_i = p_0$.

Now, we are mainly concerned here with the Bayesian setting after termination of the algorithm, where the estimated failure probability $\hat{p}_f$ is given and the thresholds are also deterministic. Anyway, although thresholds are deterministic after termination of the algorithm, we still have to deal with the structure of creation of the subsets where they are stochastic at first. This defines the distribution of the intermediate probabilities.

Definition 4.2.5 (Stochastic Variable: Limit State Value)**.** *Define* $G := g(\mathbf{X})$, *the real valued random variable given by the limit state value of the stochastic properties* $\mathbf{X}$. *The truncated versions of* G *corresponding to the subsets* $i = 1, ..., m$, *are written as* $G_i := (G|G \leq b_i)$. *The corresponding cdfs are referred to by* $F_{G_1}, ..., F_{G_m}$.

With Definition 4.2.5, the intermediate subset probabilities $P_{r,i} = P(\mathbb{F}_i|\mathbb{F}_{i-1})$, $i = 1, ..., m$ can be rewritten as

$$P_{r,i} = P(G_{i-1} \leq b_i) = F_{G_{i-1}}(b_i) \ .$$

The following assumption is not really restrictive in practice, but is necessary to allow inversion of the cdf of G which will be required later.

Assumption 4.2.6. *The cdf* F_G *is assumed to be strictly increasing with respect to* $g(x)$.

Although the distribution of the subset samples generated according to Definition 4.2.3 presumably follows the stationary distribution of the subsets, a thorough debate to remove remaining uncertainties is still pending. In the following, we add some arguments to support the statement, but some skepticism remains about the following remarks (Remark 4.2.7, Remark 4.2.8 and Remark 4.2.9), in particular with respect to ignoring the samples that have equality to the threshold.

Remark 4.2.7 (Hitting the Threshold)**.** *In step 1) of Definition 4.2.3, we only take the $(p_0N-\zeta)$-th lowest values $x^{r,i}_{(1)},...,x^{r,i}_{(p_0N-\zeta)}$ as seeds for the next subset. This aims at generally avoiding to hit the threshold b_{i+1}, which the limit state value $g^{r,i}_{(p_0N)} = b_{i+1}$ of $x^{r,i}_{(p_0N)}$ does with probability one by definition. In practice, there might be many samples equal to the threshold due to non-perfect sampling. Then, all such samples or even all samples of the corresponding chain highly correlated to it should not be used as seeds or samples for the next subset. Otherwise, the assumption of a stationary distribution can not hold.*

Hitting the threshold might often be negligible, but for explicit results and especially small N or high p_0 with many subsets, we should take care of such approximations. In real simulations, doubles of samples are almost certain and more dependencies are generally present by MCMC, amplifying the effect clearly. It is also important that the bias by keeping threshold hitting values is negative so that it might result in an underestimation of the failure probability. The samples hitting the threshold slow down approaching the final threshold b^* as they are concentrated most far from the failure region of possible sample outcomes. Of course removing samples has to be done carefully also, as the seed number and thus often the chain number is reduced. A high enough chain number, however, is necessary to deal with possible ergodicity issues of the algorithm (compare Remark 3.1.14).

Remark 4.2.8 (Stationary Distribution and Dropping Samples)**.** *By Remark 3.1.15 samples produced in 1) are in the stationary distribution of $\mathbb{F}_i$ from start. This is easy to see by the definition of the subsets and then follows by the symmetric proposal distribution in Markov chain Monte Carlo for higher level subsets. However, we have to be careful as it is clear that Remark 3.1.15 can only hold if we drop the threshold hitting sample and do not use it for estimation of further intermediate subset probabilities.*

Next, we would like to consider the stationarity of the samples in subset simulation in a more formal and detailed way. In this way, we aim above all at detecting issues that could still cause problems.

Remark 4.2.9 (Stationary Distribution: More Formal Considerations)**.** *It seems that the algorithm in Definition 4.2.3, if necessary supplemented by the extension of also not using related samples in general (as indicated in Remark 4.2.7), should lead to samples in the stationary distribution for each*

subset, i.e. samples in each subset are distributed identically as G_i, $i = 0, 1, ..., m$. *In the remainder of this remark, we address this claim and investigate to what extent a formal proof could be obtained.*

First, it is reasonable to prove the claim by induction over the subset level i. *In the first step of Subset Simulation, we produce* N *Monte Carlo samples with limit state values* $g_1^{r,0}, ..., g_N^{r,0}$ *distributed according to* G. *Then, the first threshold is defined as* $b_1 := g_{(p_0 N)}^{r,0}$. *Accordingly, the first subset is defined as* $\mathbb{F}_1 = \{\mathbf{x} \in D | g(\mathbf{x}) \leq b_1\}$. *Now, as proposed in Remark 4.2.8, the samples* $x_{(1)}^{r,0}, ..., x_{(p_0 N - \zeta)}^{r,0}$ *are used as seeds for* $\mathbb{F}_1$, *while we do not further use the samples with limit state value equal to the threshold of subset* $\mathbb{F}_1$ *as these were used to define the new subset. Then the remaining samples are distributed according to* $G_1 = (G | G \leq b_1)$ *by definition. Intuitively we might think that dropping the samples defining the threshold, which are also the samples with highest limit state value in the subset by definition, might lead to a bias. However, this is not the case as these samples, considered with perspective of the new subset, only correspond to a random selection of the threshold, which should be performed independently of the other samples, used as seeds. Thus, the only difficulty might come from dependencies between samples in higher subset levels. Yet, the effect of the correlated samples should be a minor one, in particular if the proposal spread is high enough as this leads to doubled samples rather than highly dependent ones with differing limit state values. The next* $(1 - p_0)N + \zeta$ *samples that are produced by MCMC are also in the stationary distribution by Remark 3.1.12. Repetitive application of the above argument would yield the claim for all subsets when neglecting or circumventing the discussed difficulty regarding the threshold sample and upcoming correlations in higher order subsets by MCMC.*

4.2.2 Statistical Analysis and Unbiased Estimation

Based on the interpretation by order statistics, we may develop explicit formulas for bias and variance of the Subset Simulation estimator, allowing to extend past asymptotic results also to small sample numbers and even allowing to decrease bias and variance in general. In particular, the bias correction as well as the stochastic interpretation, which will be derived in this section, do help in practical application of the algorithm, but also allow to decrypt effects that could not yet be identified.

Before proceeding with derivation of the distribution of the true interme-

diate subset probabilities, we first need to specify some notation.

Definition 4.2.10 (Beta Distribution)**.** *The Beta distribution* $\text{Beta}(\alpha, \beta)$ *with parameters* $\alpha > 0$ *and* $\beta > 0$ *is given by the pdf*

$$f_\beta(p; \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$

for $p \in [0, 1]$*.* $B(\alpha, \beta)$ *is the Beta function*[1]*.*

Proposition 4.2.11 (Distribution of the True Intermediate Subset Probabilities)**.** *Under Assumption 3.1.10, constructing subsets according to Definition 4.2.3 and if there would be no bias from the stopping criterion (compare Remark 4.2.13), the true intermediate subset probabilities* $P_{r,i} = P(\mathbb{F}_i|\mathbb{F}_{i-1})$*,* $i = 1, ..., m-1$ *follow a Beta distribution* $\text{Beta}(\alpha, \beta)$ *with parameters* $\alpha = p_0 N$ *and* $\beta = (1 - p_0)N + 1$*, i.e.*

$$P_{r,i} \sim \text{Beta}(p_0 N, (1 - p_0)N + 1) \ .$$

Proof. Let $i = 0, 1..., m-2$ be arbitrary[2]. We start with the definition, using the truncated versions G_i of $G = g(\mathbf{X})$ as in Definition 4.2.5, allowing us to rewrite the true intermediate subset probability as

$$P_{r,i+1} = P(G_i \leq b_{i+1}) \ .$$

Reformulation according to the subset construction in Definition 4.2.3, defining b_{i+1} successively as the $p_0 N$-th order statistic of the current subset, yields

$$P(G_i \leq b_{i+1}) = P(G_i \leq g^{r,i}_{(p_0 N)}) \ .$$

Now, $x^{r,i}_1, ..., x^{r,i}_N$ are iid by Assumption 3.1.10 and follow the stationary distribution of subset $\mathbb{F}_i$ by Remark 3.1.15. The corresponding limit state values $g^{r,i}_1, ..., g^{r,i}_N$ are then distributed according to the corresponding truncated distribution F_{G_i} of G. Moreover, using that for a standard uniform distribution $U \sim \mathcal{U}(0, 1)$ and an arbitrary random variable with strictly increasing cdf, such as G_i, $F^{-1}_{G_i}(U) \sim G_i$ holds, we can conclude

$$P(G_i \leq g_i^{(p_0 N)}) = P(G_i \leq F^{-1}_{G_i}(u^{r,i}_{(p_0 N)}))$$

[1] More details are given in Appendix B.

[2] The shifted index simplifies notation.

where $u_1^{r,i}, ..., u_N^{r,i}$ are samples drawn from a standard uniform distribution $U \sim \mathcal{U}(0,1)$ and $u_{(1)}^{r,i}, ..., u_{(N)}^{r,i}$ their corresponding order statistics. Using the definition of the cdf yields

$$P(G_i \leq F_{G_i}^{-1}(u_{(p_0N)}^{r,i})) = F_{G_i}(F_{G_i}^{-1}(u_{(p_0N)}^{r,i})) = u_{(p_0N)}^{r,i} \ .$$

This yields

$$P_{r,i+1} = u_{(p_0N)}^{r,i}$$

for an arbitrary $i \in \{0, 1, ..., m-2\}$. Lastly, known for order statistics, the distribution of $u_{(p_0N)}^{r,i}$ is a Beta distribution $\text{Beta}(\alpha, \beta)$ with parameters $\alpha = p_0N$ and $\beta = (1-p_0)N + 1$ (cf. Arnold et al. (1992)).

□

The last subset $\mathbb{F}_m$ has a special role in Subset Simulation (compare Definition 4.2.3). Instead of being adaptively created, it has a fixed threshold $b_m = b^*$. If we do not fix $\hat{p}_f$ in advance, $\hat{p}_m$ is stochastic in contrast to $\hat{p}_i = p_0$, $i = 1, ..., m-1$. Still, under the assumption of given $\hat{p}_m$, we however get information directly by the outcome of $P_{r,m}$ instead. In comparison to the other subsets, it is appropriate to use a standard Bayesian approach instead of order statistics (compare Remark 4.2.31).

Remark 4.2.12 (Distribution of the Last Subset's Intermediate Probability)**.** *The intermediate probability of the last subset $P_{r,m}$ is Beta distributed*

$$P_{r,m} \sim \text{Beta}(\hat{p}_m N, (1-\hat{p}_m)N) \ .$$

This follows by Bayes statistics under choice of a Beta distributed non-informative prior with $\alpha = 0$, $\beta = 0$ (Haldane prior) for $P_{r,m}$ and the Binomial distributed data likelihood for updating given by the outcome of the Monte Carlo simulation. Such a prior is admissible, since we have guaranteed successes and failures in the Bernoulli trials by definition of the subset. Compare Tuyl et al. (2008) for more details on non-informative priors and note that the drawbacks stated for the Haldane prior are excluded by construction of the subsets in Subset Simulation, supporting the proposed choice. Anyhow, we neglect the fact that the last subset is generated as a stopping criterion, meaning that our non-informative prior is not the perfect choice but remains a plausible one since this effect should play a minor role in general.

Based on the simplification by a non-informative prior, similar effects occur in the case of the other subsets.

Remark 4.2.13 (Distribution Changes by Stopping Criterion)**.** *The condition in Proposition 4.2.11 does not hold exactly. Similar to Remark 4.2.12, we had truncated Beta distributions for all subsets $\mathbb{F}_i, i = 1, ..., m-1$. The knowledge of m does indeed fix all previous intermediate probabilities $P_{r,i}$, $i = 1, ..., m-1$ to a value greater than $P_{r,i} \geq \frac{p_f}{\prod_{j=1}^{i-1} P_{r,j}}$. This also introduces a negative bias in estimation, thus an underestimation of the failure probability, where its extent increases in higher subset levels.*

The effects of the stopping criterion will be neglected in the following, until its inclusion will be discussed in Section 4.2.4, Remark 4.2.33. Impacts on the result of Subset Simulation are often negligible, especially when $\frac{\log(p_f)}{\log(p_0)}$ is not approximately an integer and N is not very small. In a next step, we devote ourselves to more general conditions, showing independence between successive intermediate subset probabilities under weak assumptions. This is in particular necessary to derive a formula for the bias but also an important discovery on its own.

Remark 4.2.14 (Dependencies by MCMC)**.** *The dependency structure of samples generated by MCMC in subset $\mathbb{F}_i$ is not significantly related to typical variations in its threshold b_i. This means that although the dependencies by MCMC sampling generally increase for lower subset probabilities and thus with increasing subset level, slightly varied subset probabilities will not result in a noticeable change. A good reason for this claim to hold is the irrelevance of the subset threshold for the Markov chain creation procedure, especially in adaptive MCMC sampling (compare I3-I6). Also correlations are tried to be kept similar by adaptive MCMC for subset levels of higher order so that independence should hold.*

As a result, also the dependency structures of MCMC samples in subset $\mathbb{F}_i, i = 2, ..., m$ will not rely on previous realizations $p_{r,1}, ..., p_{r,i-1}$ if the dependency structure is not generally changing for different intermediate probabilities in $p_{r,i-1}$. Then the seeds that will be used for sampling in the next subset will have different dependencies, resulting in dependency of samples in $\mathbb{F}_i$ by MCMC with seeds in $\mathbb{F}_{i-1}$ on $P_{r,i-1}$. By Proposition 4.2.11, the distribution of the intermediate failure probabilities is given by $\text{Beta}(p_0 N, (1-p_0)N + 1)$*. It is easy to see, that this distribution depends on the number of effective samples, considering dependencies by MCMC. Thus the intermediate probability $P_{r,i-1}$ can lead to a change in distribution of samples in $\mathbb{F}_i$. However, generally this effect will only be relevant for extreme cases and simulations*

suggest that under appropriate sampling it is negligible.

Furthermore, a critical case would occur if for example $p_{r,1}, ..., p_{r,i-1}$ are exceptionally small realizations so that $\mathbb{F}_i$ has an exceptionally low threshold and samples will be more dependent, changing the dependency structure. This however is an extreme case again and it is also not clear how to compare different algorithm outcomes rigorously as the total number of subsets m is also likely to change for extreme realizations, resulting in a questionable interpretation of changes. So, although not strictly correct, it is reasonable to assume dependencies of MCMC samples in a subset to not rely on other intermediate probabilities. This does not at all induce independent samples in successive subsets. When re-using seeds, we even have repeated samples, but the exact realization does not influence the results of both intermediate subset probabilities.

Concluding, it is plausible to assume dependencies of the samples by MCMC in subset $\mathbb{F}_j, j \in \{1, ..., m\}$ independent of $P_{r,i}, i = 1, ..., m$. We conjecture that this feature might have had too much weight in past studies, utilized for capturing unknowns that have not been exposed yet. For small p_0, simulations suggested that the coefficient of variation of Subset Simulation can be approximated well when intermediate subset probabilities (Remark 4.2.17) are assumed independent. There are good reasons that independence holds under independence of the dependency structure of MCMC samples with respect to the intermediate subset probabilities, thus if the dependencies are as proposed in this remark.

We have several seeds starting from a previous subset. As those partially keep their dependency in the previous subset and higher dependencies introduce a higher bias for the intermediate subset probability, the dependency structure that is given in the seeds might propagate. Nevertheless, this is often negligible and can be further reduced by an additional randomized seed selection step. This additional step goes as follows. First drop threshold related seeds so that we are in the stationary distribution. Then additionally randomly only pick a specific amount of seeds that will be used further and drop the others completely. For example for $N = 1000, p_0 = 0.1$ we had $N_c = 100$ seeds that would create chains. Now, only picking e.g. 20 seeds and creating $N_c = 20$ chains, would then 'destroy' given dependencies between the samples, thus the dependency structure of samples becomes closer to full independence of previous intermediate subset probabilities.

Remark 4.2.15 (Destruction of Dependency Structure)**.** *The dependency structure of MCMC seeds in Subset Simulation can be destroyed by random selection of a fraction of the available seeds and not using the others for evaluation of the next subset at all.*

This will also be in favor of the independence assumption between Markov chains for analysis. On the other hand, it will increase the necessary evaluations if seed samples are kept, which is actually recommended. In our short example, we would have $100 - 20 = 80$ additional evaluations, thus $0.08N$. In total, the difference is small and it even might be sufficient for our purpose to drop less seeds for chain creation, giving the opportunity to increase efficiency. More care has to be taken with respect to ergodicity considerations (see Remark 3.1.14).

Proposition 4.2.16 (Independence of Intermediate Subset Probabilities)**.** *If Remark 4.2.14 holds or Assumption 3.1.10 is fulfilled, the intermediate subset probabilities in Subset Simulation are pairwise independent for successive subsets, i.e.*

$$P(P_{r,i+1} \leq p^{\circ} | P_{r,i} = p^{\bullet}) = P(P_{r,i+1} \leq p^{\circ})$$

for arbitrary $p^{\circ}, p^{\bullet} \in (0,1)$ *and* $i = 2, ..., m-1$*. The last subset probability is excluded.*

Proof. We prove the claim by induction on the subset level i. Starting with $i = 2$, we first analyze the structure of the first subset. We are given the sorted samples $x^{r,0}_{(1)}, ..., x^{r,0}_{(N)}$, produced by standard Monte Carlo, with limit state values $g^{r,0}_{(1)}, ..., g^{r,0}_{(N)}$ and the corresponding threshold of the first subset, defined as $b_1 = g^{r,0}_{(p_0 N)}$. Now $P_{r,1} = p^{\bullet}$ means that the threshold b_1 defines a subset with real probability $p^{\bullet}$, i.e.

$$P_{r,1} = F_G(b_1) = P(g(\mathbf{X}) < b_1) = p^{\bullet} \ .$$

By Definition 4.2.3 and Remark 3.1.15, the seeds $g^{r,1}_{(1)}, ..., g^{r,1}_{(p_0 N - 1)}$ of the next subset are already in its stationary distribution and samples by MCMC are distributed accordingly. Thus, the limit state value $g^{r,1}_s$ of a randomly chosen sample of the next subset $x^{r,1}_s, s \in \{1, ..., N\}$ fulfills

$$F_G(g^{r,1}_s) \sim (U | U < p^{\bullet}) \sim \mathcal{U}(0, p^{\bullet}) \tag{4.5}$$

with $U \sim \mathcal{U}(0,1)$ the standard uniform distribution, $(U|U < p^\bullet)$ the corresponding truncated distribution and $\mathcal{U}(0,p^\bullet)$ the uniform distribution with pdf constant $\frac{1}{p^\bullet}$ on $(0,p^\bullet)$. For an arbitrary $p^\circ \in (0,1)$ we have

$$P\left(\frac{F_G(g_s^{r,1})}{F_G(g_{(p_0N)}^{r,0})} \leq p^\circ \middle| F_G(g_{(p_0N)}^{r,0}) = p^\bullet\right) = P\left(\frac{F_G(g_s^{r,1})}{p^\bullet} \leq p^\circ\right) \ .$$

Using $\frac{F_G(g_s^{r,1})}{p^\bullet} \sim U$ by equation (4.5) results in

$$P(U \leq p^\circ) = p^\circ$$

for all p° for an arbitrary sample $x_s^{r,1}$ which means

$$\left(P_{r,2}|P_{r,1} = p^\bullet\right)$$

follows a standard uniform distribution independent of the previous subset's conditional probability $p^\bullet$. This shows independence of the single sample intermediate distribution with respect to the previous subset. As independence holds for an arbitrary sample, it remains to show that it also holds for the order statistics of several such samples even under dependencies between those samples. First, we proceed the same way as in the single sample case, having

$$\begin{aligned} P(P_{r,2} \leq p^\circ|P_{r,1} = p^\bullet) &= P\left(\frac{F_G(g_{(p_0N)}^{r,1})}{F_G(g_{(p_0N)}^{r,0})} \leq p^\circ \middle| F_G(g_{(p_0N)}^{r,0}) = p^\bullet\right) \\ &= P\left(F_G(g_{(p_0N)}^{r,1}) \leq p^\circ p^\bullet\right) \ . \end{aligned} \tag{4.6}$$

This time however, we have to represent the cdf of an order statistics $F_G(g_{(p_0N)}^{r,1})$ by an equivalent order statistics $u_{(p_0N)}^{r,1}$ of N samples $u_1^{r,1}, ..., u_N^{r,1}$ drawn from a properly chosen uniform distribution with limits $(0,p^\bullet)$. Keeping the dependency structure, we may scale those variables to the interval $(0,1)$, adding the scaling factor $p^\bullet$. This leads to the same distribution for the p_0N-th order statistics $p^\bullet u_{(p_0N)}^{r,1*}$ of $p^\bullet u_1^{r,1*}, ..., p^\bullet u_N^{r,1*}$ with $u_j^{r,1*}$, $j = 1, ..., N$ the corresponding variables on the interval $(0,1)$. Then (4.6) becomes

$$P\left(u_{(p_0N)}^{r,1*} \leq p^\circ\right)$$

similarly as above in the single sample case, but relying on an order statistic of uniform distributed samples. The samples are explicitly allowed to be dependent due to the fact that they have to cover the dependencies induced by Markov Chain Monte Carlo. Now this distribution is independent of the previous subset if the dependencies between the samples $u_1^{r,1*}, ..., u_N^{r,1*}$ do not depend on the previous subset probabilities. Thus Remark 4.2.14 yields the claim.

As we are in the stationary distribution for crude Monte Carlo and Markov chain Monte Carlo and also Remark 4.2.14 does not depend on the subset level, our argument can be applied successively also on higher order subsets to prove the claim.

□

Proposition 4.2.16 yields independence of successive intermediate probabilities. However, this will not suffice our demands, and there are good reasons for an extension of the result to a more general case.

Remark 4.2.17. *As the subsets are created successively and only seem to rely on the last subsets by the subset seeds received from the previous subset, it is plausible to assume that also mutual independence of intermediate probabilities holds as a consequence of Proposition 4.2.16.*

Evidence for independence was found in simulations (also compare Section 4.2.3). Thus, we assume Remark 4.2.17 generally holds in our theoretical studies. For low p_0, it is important to destroy the dependency structure as it is then more likely to have extreme cases, as described in Remark 4.2.14, due to the small number of samples considered for the order statistics. As already mentioned earlier, this can be done by selecting only a few of the available samples for further evaluation.

Remark 4.2.18. *Independence of the intermediate subset probabilities also guarantees independence of the error in estimation of the individual subset probabilities. It is also worthwhile that we did not require independence of within subset samples in Remark 4.2.17 so that it holds for Subset Simulation with samples by MCMC, if Remark 4.2.14 holds. This later allows to switch more easily from analysis under independence to realistic settings considering dependencies between samples.*

Following the approach in Zuev et al. (2012), we also use the following approximation, given in Fan (1991), to derive the posterior distribution of P_f. Note, we use the notation corresponding to Subset Simulation.

Theorem 4.2.19. *For $P_{r,1}, P_{r,2}, ..., P_{r,m}$ independent Beta variables, $P_{r,i} \sim \text{Beta}(\alpha_i, \beta_i)$ their product $P_f = P_{r,1}P_{r,2} \cdot ... \cdot P_{r,m}$ is approximately Beta distributed* $\text{Beta} \sim (\alpha, \beta)$ *with*

$$\alpha = \mu_1 \frac{\mu_1 - \mu_2}{\mu_2 - \mu_1^2},\ \beta = (1 - \mu_1)\frac{\mu_1 - \mu_2}{\mu_2 - \mu_1^2}$$

where

$$\mu_1 = E[P_f] = \prod_{i=1}^{m} \frac{\alpha_i}{\alpha_i + \beta_i} \text{ and } \mu_2 = E[(P_f)^2] = \prod_{i=1}^{m} \frac{\alpha_i(\alpha_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)}\ .$$

Accurateness of the approximation is also discussed in Fan (1991). Similar to Zuev et al. (2012), we can now derive a distribution for the true failure probability given a Subset Simulation estimate. The procedure and result are exactly the same, except for having different distributions for intermediate subset probabilities and not needing independence as an unfulfilled assumption but instead having good reasons for independence by Remark 4.2.17 so that we can have more confidence in the outcome.

Corollary 4.2.20. *[Posterior of the True Failure Probability] For ease of notation and computation, assume $\hat{p}_f = p_0^m$. Then after termination of Subset Simulation, given parameters p_0, N, the posterior of the true failure probability P_f is approximately Beta distributed* $\text{Beta}(\alpha_{\hat{p}_f}, \beta_{\hat{p}_f})$ *with parameters*

$$\alpha_{\hat{p}_f} = \frac{\left(\frac{p_0 N}{N+1}\right)^m \left(1 - \left(\frac{p_0 N+1}{N+2}\right)^m\right)}{\left(\frac{p_0 N+1}{N+2}\right)^m - \left(\frac{p_0 N}{N+1}\right)^m}$$

and

$$\beta_{\hat{p}_f} = \frac{\left(1 - \left(\frac{p_0 N}{N+1}\right)^m\right)\left(1 - \left(\frac{p_0 N+1}{N+2}\right)^m\right)}{\left(\frac{p_0 N+1}{N+2}\right)^m - \left(\frac{p_0 N}{N+1}\right)^m}\ .$$

Proof. This is a direct consequence of Theorem 4.2.19, plugging in the distribution of the intermediate subset probabilities as given in Proposition 4.2.11. Also note that independence was used, between subsets (Remark 4.2.17) as well as within subsets (Assumption 3.1.10). The intermediate distribution parameters are given by $(\alpha_i, \beta_i) = (p_0 N, (1 - p_0)N + 1)$, $i = 1, ..., m - 1$ and $(\alpha_m, \beta_m) = (\hat{p}_m N, (1 - \hat{p}_m)N)$ by Proposition 4.2.11 and Remark 4.2.12. □

We are not yet able to apply the result in Corollary 4.2.20 as the number of samples within the subsets is assumed independent. A subsequent discussion on the effective sample number follows later, starting with Remark 4.2.30. Next, we want to derive a formula that allows to correct biased Subset Simulation results to unbiased ones. To do so, it is necessary to examine the probability of the case $\alpha_{\hat{p}_f} < 1$ in the Frequentist setting, later allowing us to proof unbiasedness for an altered Subset Simulation estimator under usage of this conclusion. For that reason, we look at the relation between estimated failure probability $\hat{p}_f$ and a resulting $\alpha_{\hat{p}_f} < 1$.

Remark 4.2.21 (Relation $\hat{p}_f$ and $\alpha_{\hat{p}_f} < 1$)**.** *Using Corollary 4.2.20, but also allowing $\hat{p}_f \neq p_0^m$ (derivation by Theorem 4.2.19 is straightforward), we can take a numerical approach. Given an estimated failure probability $\hat{p}_f$ and the Subset Simulation parameters p_0 and N, we have a unique suitable number of subsets m and the estimated probability in the last subset $\hat{p}_m$. As a consequence, all required information to derive the corresponding Beta distribution* $\text{Beta}(\alpha_{\hat{p}_f}, \beta_{\hat{p}_f})$ *is given. It is thus straightforward to check for which parameters p_0 and N, a given $\hat{p}_f$ leads to $\alpha_{\hat{p}_f} < 1$. The result is best captured through illustration. Figure 4.7 includes most relevant constellations.*

Figure 4.7 allows to relate the requirements on N, given Subset Simulation with parameter p_0, to have an acceptable probability for $\hat{p}_f$ yielding $\alpha_{\hat{p}_f} < 1$. It remains to estimate the probability of critical outcomes $\hat{p}_f$, given p_f, in the Frequentist setting. This estimation does not need to be very accurate so that it is not necessary to know the probability distribution of $\hat{P}_f$, given p_f, for our purposes. Instead, we know that the shape typically (as a product of independent stochastic variables) yields a heavy tailed distribution, typically well suiting a lognormal distribution (compare Breitung (2019)). Furthermore, the coefficient of variation can be bounded to realistic cases given that the sample number N is above some limit. The boundaries do not need to be tight, they should rather provide a good heuristic argument and serve as a basis to set safety margins for derivation of approximate results.

To get an intuition of such distributions $\hat{P}_f$, we next give an example.

Example 4.2.22 (How Likely is $\alpha_{\hat{p}_f} < 1$?)**.** *For demonstration, we choose $N = 200$, $p_0 = 0.1$ and apply* 1000 *independent simulation runs of Subset Simulation. We thereby aim at tracing variation of the estimated failure probabilities on the downside to understand which N values in Figure 4.7*

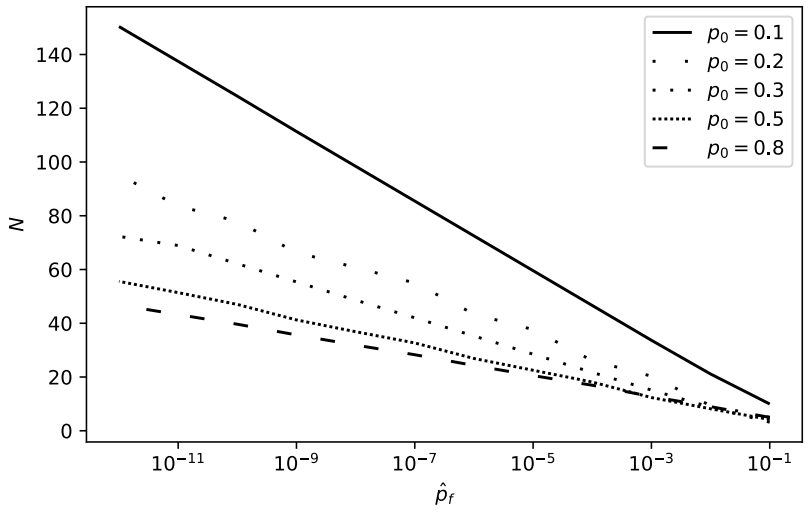


Figure 4.7: Relation $\hat{p}_f$ and $\alpha_{\hat{p}_f} < 1$. The minimal (effective) sample number N per subset which leads to an $\alpha_{\hat{p}_f} < 1$ of the Beta distribution in Corollary 4.2.20, given a specific estimated failure probability $\hat{p}_f$ by Subset Simulation with parameter p_0, is shown.

become relevant for specific true failure probabilities. In the example, we fix $p_f = 1 \cdot 10^{-6}$*. Figure 4.8 illustrates the relevance of estimated failure probabilities. The* 0.025*-quantile was found to be at around* 10% *of the original failure probability, and outliers, except for one, took values above* 5% *of* $p_f = 1 \cdot 10^{-6}$*. The average empirical correlation factor over all independent runs was found below two, so that we assume an effective sample number higher than* $N = 100$*, of course under the assumption that autocorrelations sufficiently well reflect the dependencies in the corresponding order statistics. Figure 4.7 then allows to identify the estimates which would yield an* $\alpha_{\hat{p}_f} < 1$*, concluding that such estimation results are very unlikely.*

Simulations suggest that the findings in Example 4.2.22 should also approximately hold for other true failure probabilities and other choices of p_0. Indeed we may conclude that estimated failure probabilities below 5% of the true failure probability should be very unlikely for many cases, in particular for $1 \cdot 10^{-8} < p_f < 1$. Other sample numbers per subset N of course result in different variations of the estimated failure probabilities. However, for higher N, variation decreases so that we are on the safe side and for lower N we are

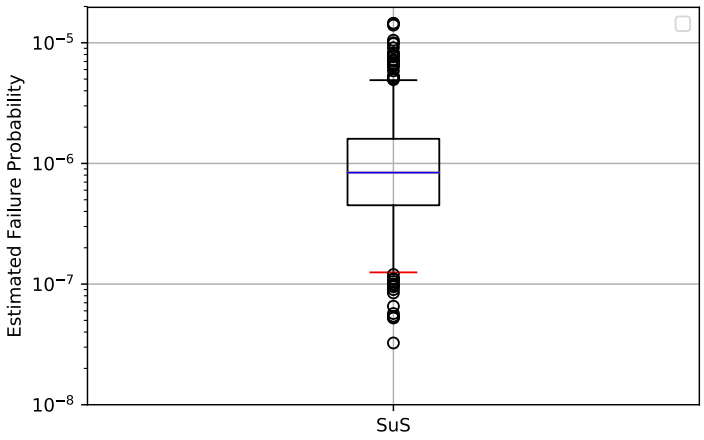


Figure 4.8: Downside Probabilities. The distribution of estimated failure probabilities for $N = 200$, $p_0 = 0.1$, $p_f = 1 \cdot 10^{-6}$ given 1000 independent simulation runs of Subset Simulation. Outliers take values down to $5 \cdot 10^{-8}$ and the whiskers were chosen to contain 95% of the values, providing a downside 2.5%-quantile slightly above $1 \cdot 10^{-7}$.

in a questionable situation anyways (e.g. with respect to critical ergodicity issues).

Remark 4.2.23 (Probability of $\alpha_{\hat{p}_f} < 1$)**.** *In the Frequentist setting and under assumption of independent within subset samples (Assumption 3.1.10), given p_f and applying Subset Simulation with parameters p_0 and sample number N, we may approximate the probability of $\alpha_{\hat{p}_f} < 1$ as follows. For a given p_f and p_0, we use Figure 4.7 to find the corresponding N value that belongs to e.g. $\hat{p}_f := 0.05 p_f$. We then need to add a safety margin by deriving a non-effective sample number required to achieve this number of effective samples N. For example a factor of $2-4$ covers most settings in modern implementations (I3,I4), dependent on p_f and most likely also p_0. Evaluation of the corresponding effective N after finished simulation can be approached as in Assumption 3.1.16 and Section 4.2.3. Concluding, it is easy to choose N so that the set of $\hat{p}_f$ yielding an $\alpha_{\hat{p}_f} < 1$ becomes negligible, since the initial values for comparison as in Figure 4.7 are very low already.*

Although we need to require a minimum N under some uncertainties here, the unknowns can be faced by boundary approaches or safety margins and we receive explicit advisory instead of asymptotic results, stating explicit

values for N. Simulations in Section 4.2.3 below also provide evidence for the thereby drawn conclusions.

The Bayes framework itself can be useful to derive confidence intervals for the true failure probability. However, a main focus in research is also the optimization of the algorithm, deriving good choices for p_0, or as shown in the following, remove the bias of the result. This, on the one hand, relies on the Frequentist approach and requires a fixed p_f. On the other hand, we do not have the distribution of $\hat{P}_f$ for a given p_f. So we need an argument for symmetry by the Bayes Formula as provided. To do so, it is however necessary to first recognize that $P_f|\hat{p}_f$ is a continuous random variable taking values in $(0,1)$, whereas $\hat{P}_f|p_f$ is a discrete variable taking values in $\{\hat{p}_{f,i} : \hat{p}_{f,i} \text{ is a possible outcome of Subset Simulation}\} =: A_{\text{SuS}}$. We reference its corresponding probability mass function with $\mathbf{p}_{\hat{P}_f}$. This never defines a continuous variable as N is practically finite always. Without loss of generality assume that A_{SuS} is a finite set with n_p elements and let the possible outcomes $\hat{p}_{f,1} < \hat{p}_{f,2} < .... < \hat{p}_{f,n_p} = 1$ be given in ascending order. To apply Bayes Formula, we need to assume a prior, where we aim to take a non-informative one. For P_f, we assume a uniform prior $\mathcal{U}(0,1)$. On the other hand for $\hat{P}_f$, it would not be plausible to use the standard discrete uniform prior, since for $N \to \infty$, it would be a desirable property that the prior converges to something similar to the continuous uniform prior. We achieve this, by defining $\hat{p}_{f,0} = 0$ and then assigning the probabilities as follows:

$$\mathbf{p}_{\hat{P}_f} : A_{\text{SuS}} \to (0,1), (\hat{p}_{f,i}) \mapsto \hat{p}_{f,i} - \hat{p}_{f,i-1} \ .$$

It is easy to check that $\sum_{i=1}^{n_p} \mathbf{p}_{\hat{P}_f}(\hat{p}_{f,i}) = 1$ and that for high N in Subset Simulation, this approximates the continuous uniform probability distribution.

We then have for an arbitrary $p_f \in (0,1)$ and an arbitrary $\hat{p}_{f,i} \in A_{\text{SuS}}$ by the Bayes Formula

$$f_{P_f|\hat{P}_f=\hat{p}_{f,i}}(p_f) = \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \frac{f_{P_f}(p_f)}{\mathbf{p}_{\hat{P}_f}(\hat{p}_{f,i})} = \frac{1}{(\hat{p}_{f,i} - \hat{p}_{f,i-1})} \cdot \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \ . \tag{4.7}$$

Using this symmetry, we can now prove how to approximately remove the bias in Subset Simulation by utilizing results of the Bayes approach.

Proposition 4.2.24 (Unbiased Subset Simulation Estimation: Proof of Concept)**.** *Under the assumption that the probability of $\hat{P}_f$ leading to $\alpha_{\hat{p}_f} < 1$ is*

small (compare Remark 4.2.23), we state the following claim. Given an estimated failure probability $\hat{p}_f$ by Subset Simulation with parameters p_0, N and given m, $\hat{p}_m$, we state an approximately unbiased Subset Simulation result under fixed but unknown real failure probability p_f

$$E_{p_f}\left[\hat{P}_f^{\nu}\right] \approx p_f$$

by transforming the Subset Simulation estimation outcome $\hat{p}_f$ according to

$$\hat{p}_f^{\nu} = \hat{p}_f \left(E\left[\frac{\hat{P}_f}{P_f} \middle| \hat{P}_f = \hat{p}_f \right] \right)^{-1} .$$

The expectation used for the transformation corresponds to the Bayesian approach, referring to the distribution of the true failure probability under given estimated failure probability $\hat{p}_f$ (as stated in Definition 4.2.2).

Proof. Starting with

$$E_{p_f}\left[\hat{P}_f^{\nu}\right] = E_{p_f}\left[\hat{P}_f \left(E\left[\frac{\hat{P}_f}{P_f} \middle| \hat{P}_f \right] \right)^{-1} \right] ,$$

we use the definition of the expected value and reformulation to get[3]

$$\sum_{i=1}^{n_p} \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \left(E\left[\frac{1}{P_f} \middle| \hat{P}_f = \hat{p}_{f,i} \right] \right)^{-1} .$$

To further reformulate the term in the brackets, we apply Lemma B.0.3 (expected value of the inverse of a Beta distributed random variable) where $f_{P_f|\hat{P}_f=\hat{p}_{f,i}}$ follows a Beta distribution with parameters $(\alpha_{\hat{p}_{f,i}}, \beta_{\hat{p}_{f,i}})$, depending on $\hat{p}_{f,i}$, by Corollary 4.2.20. We thus have

$$\sum_{i=1}^{n_p} \frac{\alpha_{\hat{p}_{f,i}} - 1}{\alpha_{\hat{p}_{f,i}} + \beta_{\hat{p}_{f,i}} - 1} \cdot \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) . \tag{4.8}$$

[3]Strictly speaking, we also require an independence assumption with respect to the corresponding Frequentist and Bayesian estimators. However, since this is a plausible assumption, the derived result is an approximation and we have evidence for our claim by simulations, a detailed discussion is skipped here.

Next, by Bayes Formula (Equation 4.7) we have

$$\sum_{i=1}^{n_p} \frac{\alpha_{\hat{p}_{f,i}} - 1}{\alpha_{\hat{p}_{f,i}} + \beta_{\hat{p}_{f,i}} - 1} \cdot (\hat{p}_{f,i} - \hat{p}_{f,i-1}) \cdot f_{P_f|\hat{P}_f=\hat{p}_{f,i}}(p_f)$$

where the pdf $f_{P_f|\hat{P}_f=\hat{p}_{f,i}}$ can be identified as the pdf of the Bayes estimation of the true failure probability $f_{P|\hat{P}_f=\hat{p}_{f,i}}$ as given above. Using the definition of the pdf, this becomes

$$\sum_{i=1}^{n_p} \frac{\alpha_{\hat{p}_{f,i}} - 1}{\alpha_{\hat{p}_{f,i}} + \beta_{\hat{p}_{f,i}} - 1} \cdot \frac{p_f^{\alpha_{\hat{p}_{f,i}}-1}(1-p_f)^{\beta_{\hat{p}_{f,i}}-1}}{B(\alpha_{\hat{p}_{f,i}}, \beta_{\hat{p}_{f,i}})} \cdot (\hat{p}_{f,i} - \hat{p}_{f,i-1}) \tag{4.9}$$

with B the Beta function. In the next step, we utilize a relation of the Beta function (see Lemma B.0.2)

$$B(\alpha_{\hat{p}_{f,i}}, \beta_{\hat{p}_{f,i}}) = B(\alpha_{\hat{p}_{f,i}} - 1, \beta_{\hat{p}_{f,i}}) \cdot \frac{\alpha_{\hat{p}_{f,i}} - 1}{\alpha_{\hat{p}_{f,i}} - 1 + \beta_{\hat{p}_{f,i}}} \tag{4.10}$$

which is defined for $\alpha_{\hat{p}_{f,i}} > 1$ and undefined otherwise. As a consequence, we first need to split the integral. To do so, note that given $\hat{p}_{f,i}$ by Subset Simulation as in Remark 4.2.20, we have a unique combination of parameters $(\alpha_{\hat{p}_{f,i}}, \beta_{\hat{p}_{f,i}})$ assigned to it. Naturally, we may thus define the set of all i which correspond to non-admissible $\alpha_{\hat{p}_{f,i}} < 1$

$$A_{\text{na}} := \{i : \alpha_{\hat{p}_{f,i}} < 1\}$$

and the thereby induced decomposition of all $\hat{p}_{f,i}$

$$\{i : \hat{p}_{f,i} \in A_{\text{SuS}}\} = A_{\text{na}} \dot{\cup} A_a \ ,$$

where A_a defines the set of indexes, leading to admissible parameters for the desired transformation. Then, using (4.10) and taking the representation given in (4.8) for the non-admissible set, (4.9) becomes

$$\begin{aligned} &\sum_{i \in A_{\text{na}}} \frac{\alpha_{\hat{p}_{f,i}} - 1}{\alpha_{\hat{p}_{f,i}} + \beta_{\hat{p}_{f,i}} - 1} \cdot \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \\ +p_f &\sum_{i \in A_a} (\hat{p}_{f,i} - \hat{p}_{f,i-1}) \cdot \frac{p_f^{\alpha_{\hat{p}_{f,i}}-2}(1-p_f)^{\beta_{\hat{p}_{f,i}}-1}}{B(\alpha_{\hat{p}_{f,i}} - 1, \beta_{\hat{p}_{f,i}})} \ . \end{aligned}$$

The second term contains the pdf $\tilde{f}_{P_f|\hat{P}=\hat{p}_{f,i}}$ of a Beta distribution with parameters $(\alpha_{\hat{p}_{f,i}} - 1, \beta_{\hat{p}_{f,i}})$ and can thus be rewritten as

$$p_f \sum_{i \in A_a} (\hat{p}_{f,i} - \hat{p}_{f,i-1}) \cdot \tilde{f}_{P_f|\hat{P}=\hat{p}_{f,i}}(p_f) \ .$$

Again using Equation 4.7, then results in

$$p_f \sum_{i \in A_a} \tilde{\mathbf{p}}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \ .$$

Concluding, we have

$$\sum_{i \in A_{\text{na}}} \frac{\alpha_{\hat{p}_{f,i}} - 1}{\alpha_{\hat{p}_{f,i}} + \beta_{\hat{p}_{f,i}} - 1} \cdot \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) + p_f \sum_{i \in A_a} \tilde{\mathbf{p}}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \ .$$

It remains to bound the first summand from above, showing it is negligible and also derive for the second term in which cases $\sum_{i \in A_a} \mathbf{p}_{\hat{P}_f|P_f=p_f}(\hat{p}_{f,i}) \approx 1$. The result is then approximately unbiased. For both conclusions, it is important to understand how the sets A_a and A_{na} are connected to the Subset Simulation parameters and if there are realistic constellations that introduce a relevant A_a. Fortunately, Remark 4.2.23 yields the required information.

Note, $\frac{\alpha_{\hat{p}_{f,i}}-1}{\alpha_{\hat{p}_{f,i}}+\beta_{\hat{p}_{f,i}}-1}$ in the first summand is well approximated by the mode for small failure probabilities ($\beta_{\hat{p}_{f,i}} \gg \alpha_{\hat{p}_{f,i}}$) so that its value is close to $\hat{p}_f$. Assuming a negligible (improbable) set A_{na} (compare Remark 4.2.23), then certainly also results in a negligible summand. Thus, only the second summand remains. Again using the assumption of a small probability for A_{na} to occur (compare Remark 4.2.23) then yields the claim.

□

Proposition 4.2.24 shows, under suitable conditions, a way to transform biased Subset Simulation results to unbiased ones. The special feature of our derivation is an explicit specification of a minimum sample number, given specific Subset Simulation parameters, which will result in an approximately accurate bias correction. In previous analyses, asymptotic arguments were necessary for analysis of the bias in Subset Simulation. Nevertheless, as it is an approximation, evidence by simulations is crucial and given in Section 4.2.3. Proceeding with the derivation of an approach to receive an unbiased SuS result, it remains to explicitly state the term used for correction according to Proposition 4.2.24.

Theorem 4.2.25 (Relative SuS 'Bias')**.** *Under independent samples (Assumption 3.1.10) and termination of the algorithm as in Definition 4.2.2, the relative 'bias' in Subset Simulation is approximately (up to Remark 4.2.13) given by*

$$E\left[\frac{\hat{p}_f}{P_f}\right] = \left(\frac{p_0 N}{p_0 N - 1}\right)^{m-1} \left(\frac{\hat{p}_m N - \hat{p}_m}{\hat{p}_m N - 1}\right) \tag{4.11}$$

where m is the number of subsets.

Proof. We can rewrite $P_f = \prod_{i=1}^{m} P_{r,i}$ as the product of all true intermediate subset probabilities. Then

$$E\left[\frac{\hat{p}_f}{P_f}\right] = E\left[\frac{p_0^{m-1} \cdot \hat{p}_m}{\prod_{i=1}^{m} P_{r,i}}\right] .$$

By independence between intermediate subset probabilities (Remark 4.2.17), this can be formulated as

$$\prod_{i=1}^{m-1} E\left[\frac{p_0}{P_{r,i}}\right] \cdot E\left[\frac{\hat{p}_m}{P_{r,m}}\right] .$$

By Proposition 4.2.11, $P_{r,i}$ follows a Beta distribution $\text{Beta}(\alpha, \beta)$ with $\alpha = p_0 N$ and $\beta = (1 - p_0)N + 1$. The distribution of the last subset is given by Remark 4.2.12. Thus we can apply Lemma B.0.3 to get

$$\left(\frac{p_0 N}{p_0 N - 1}\right)^{m-1} \left(\frac{\hat{p}_m N - \hat{p}_m}{\hat{p}_m N - 1}\right) .$$

□

In accordance to Theorem 4.2.25, we can also state a bias-like term based on the outcome $\hat{p}_f$ of Subset Simulation. We refer to it as bias-like, as Bayesian estimation is generally not connected to the concept of a bias. However, the close relation to the Frequentist setting shown in Proposition 4.2.24 allows to estimate the bias of the Frequentist setting by given outcomes $\hat{p}_f$ in Subset Simulation. Thus we also refer to this term as if it was the Frequentist bias, as it converges to it when repetitively applying Subset Simulation and averaging over all estimated bias terms. Strictly speaking, it however is a random variable and not a bias.

Remark 4.2.26. *The 'bias' derived in Theorem 4.2.25 with respect to a given estimation $\hat{p}_f$ is always positive and $O(1/N)$, i.e.*

$$\left(\frac{p_0 N}{p_0 N - 1}\right)^{m-1} = \left(1 + \frac{1}{p_0 N - 1}\right)^{m-1} .$$

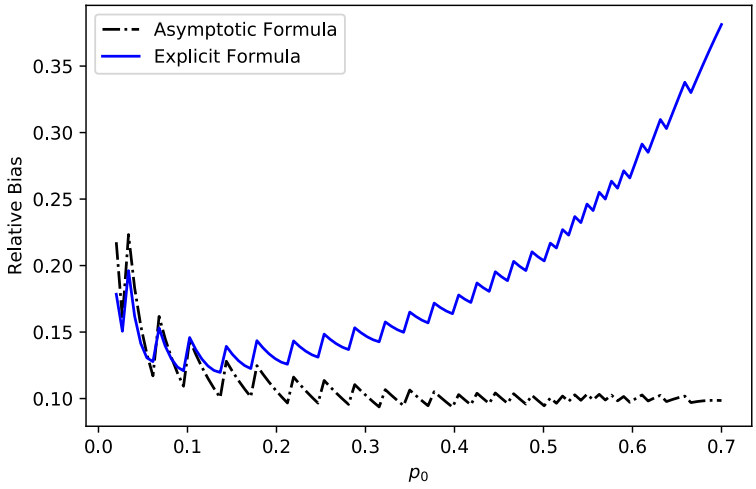
This result coincides with findings in Au and Beck (2001b) and Cérou et al. (2012).

In comparison to our result which is based on a Bayes analysis, a result under fixed $P_f = p_f$ in the Frequentist setting, requiring asymptotic arguments, was given in Cérou et al. (2012).

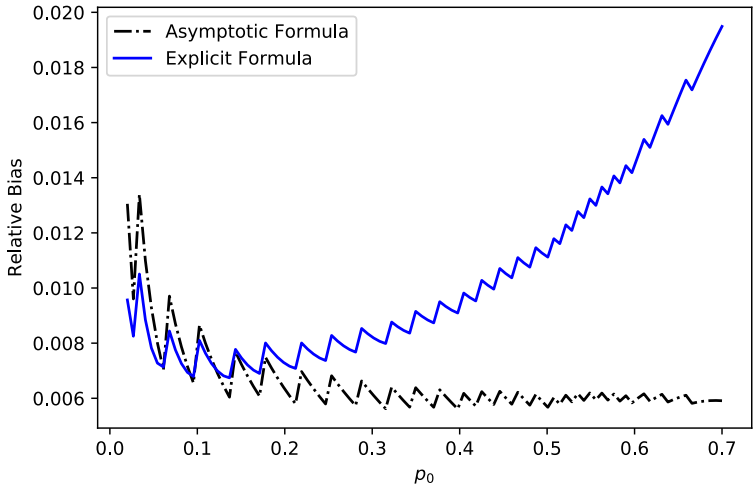
Remark 4.2.27 (Comparison of the Explicit Bias Formula to Asymptotic Results)**.** *In Cérou et al. (2012), asymptotic arguments were utilized to derive the bias in Subset Simulation as*

$$E\left[\frac{\hat{P}_f - p_f}{p_f}\right] \sim \frac{1}{N}\frac{m(1-p_0)}{p_0} . \tag{4.12}$$

A comparison of Equation 4.12 and our formula (Equation 4.11 subtracting one) is given in Figure 4.9, where for $p_{r,m} \neq p_0$ we have to adequately substitute p_0 by $p_{r,m}$ partially. It suffices to consider only these choices of E_T and p_f, since we get corresponding shapes for other constellations. For low p_0, the results appear to be similar where on the other hand high p_0, which also comes with lower N, induces a clear divergence of the results. This is a surprising result which suggests problems in derivation of one of the formulas. We believe that Equation 10 on page 807 in Cérou et al. (2012) lacks the bias in the expectation of an order statistics. This bias needs to be considered for each individual subset level, creating a divergence of the results for high numbers of subsets, i.e. high p_0 values. Anyhow, this will not affect conclusions in Cérou et al. (2012) drastically and asymptotic results remain generally valid in most points. Although the shape of the result is independent of E_T, the difference is negligible for high N_T as the bias is $O(1/N)$ in both cases. Our bias is not a prediction bias as in Cérou et al. (2012) but instead is designed to correct the bias depending on the Subset Simulation result, not requiring the value of the real failure probability. Still, the results should deliver a similar outcome for all p_0, particularly for high E_T.



(a) $E_T = 6000$



(b) $E_T = 100,000$

Figure 4.9: Comparison of Bias Formulas: The total number of evaluations $E_T = 6000$ and $E_T = 100,000$ (assumed effective number of evaluations, used in the formulas: $E_T/3$) and failure probability $p_f = 10^{-6}$ ($\hat{p}_f = 10^{-6}$ in our formula) are compared under variation of p_0. The explicit formula refers to the bias formula derived in this work where the asymptotic formula was derived in Cérou et al. (2012). Jumps are produced by a switched number of samples per subset where it is not optimal to have a high number of samples in the last subset when it has a probability close to one at the cost of reducing N also in lower level subsets with possibly low p_0.

Figure 4.9 suggests that the bias can have a significant effect, in particular for high p_0. This conclusion relies on the assumption that our formula derived in the Bayes approach also reflects bias in the Frequentist setting approximately, which is plausible. For repetitive runs and low N the bias even becomes increasingly important, due to the decreased coefficient of variation but unchanged bias. It remains to check how it suits results by simulation, which we do in Section 4.2.3. Additional to a comparison of the methods and a general bias curve, Figure 4.9 exposes spikes with respect to p_0 in the relative bias curve. This relates to the fact that we do not adapt the sample number in the last subset. So, efficiency could be increased if an adaption is performed, or if we switch to a p_0 that defines a local minimum. So, dependent on p_f, which can be estimated after a first run of Subset Simulation, we should adapt p_0 so that we choose the best p_0 and are not at a local peak or adapt the sample number in the last subset. Although generally unimpressive as we are in $O(1/N)$, this effect may lead to a for example 80% higher bias in relation to a similar p_0 at a minimum position. Unfortunately, for low N it is also harder to control the intermediate probability of the last subset because of the higher variation in general. Additionally, if we hit a p_0 that results in a high estimated intermediate probability for the last subset, the absolute value of the bias by the stopping criterion will increase (Remark 4.2.13).

Knowing an explicit formula for the bias also allows to transform the biased result to an unbiased one.

Corollary 4.2.28 (Unbiased Subset Simulation)**.** *An approximately unbiased estimation by Subset Simulation, terminated with result $\hat{p}_f$, m subset levels and $\hat{p}_m = p_m$, is given by*

$$\hat{p}_f^{\nu} = \hat{p}_f \cdot \nu$$

for $\nu = \nu_0^{m-1}\nu_m$ with $\nu_0 := \frac{p_0 N - 1}{p_0 N}$ and $\nu_m = \frac{N p_m - 1}{N p_m - p_m}$.

Proof. Follows directly by Theorem 4.2.25. □

The benefit of a bias correction is also an increase in efficiency.

Remark 4.2.29 (Improved Efficiency by Bias Correction)**.** *The bias correction as in Corollary 4.2.28 reduces the variance of the SuS estimator by canceling out the positive bias, lowering the reference value by downscaling while keeping a constant coefficient of variation.*

In the previous analysis of this section, we did not yet pay much attention to the within subset sample dependencies, mostly sticking to Assumption 3.1.10. Such dependencies and the thereby resulting effective sample number of course need to be considered in the given formulas. Unfortunately, this is not straightforward, as Section 4.1 already suggests.

Remark 4.2.30 (Correlated Within Subset Samples: Empirical Approach)**.** *In contrast to Assumption 3.1.10, samples within subsets are dependent as they are created by MCMC. We even have a targeted acceptance rate of* 0.44 *for good efficiency of the algorithm which corresponds to about half of the samples being doubles of other samples (compare Section 4.1). Analysis of order statistics with dependent samples is a challenging topic and dependency in order statistics goes beyond Markovian dependence (Arnold et al. (1992)). Indications that theoretical statements are difficult here are also given in Section 4.1. Despite the difficulties for theoretical derivation of the effective sample number in order statistics, we can still empirically approach it as usual and see if there is enough evidence by simulations. Section 4.2.3 shows that the standard approach yields good results in case of our bias formula but does not apply to the formula for the coefficient of variation. This is not very surprising since the causes of the bias are rather simple and variation of dependent order statistics is much harder to understand.*

4.2.3 Simulation Study

This section aims at providing evidence and limitations of the theoretical results of the previous section. In detail, we examine

- Independence of intermediate subset probabilities (Proposition 4.2.16, Remark 4.2.17)
- The effect of keeping samples equal to the threshold (Remark 4.2.7)
- The bias correction formula (Corollary 4.2.28)

Additionally, we also add some information on the distribution of SuS results (compare Section 3.2). Simulations are with representative p_0 values, $p_0 = 0.1$ representing low p_0 choices with few subsets and $p_0 = 0.5$ high ones with many subsets in comparison to current implementations. This reduces computational effort of the study significantly, makes the overview

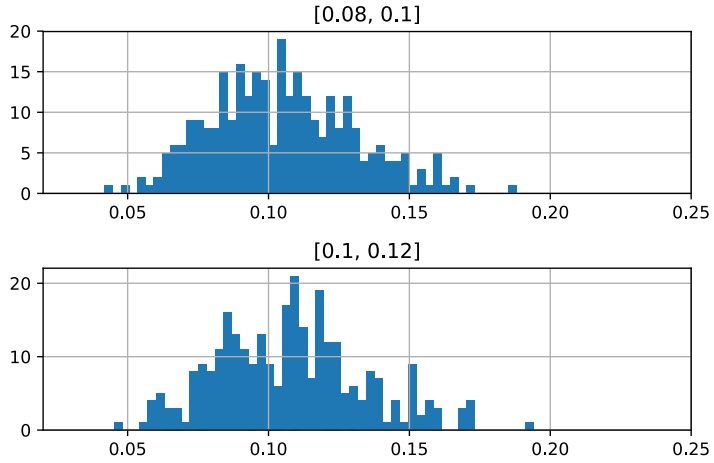
of the analysis easier manageable and still allows to answer a wide range of questions.

First, we give evidence for independence of intermediate subset probabilities by examining correlations as well as corresponding distributions. This will not let us conclude independence clearly, but at least allows us to claim that it is plausible. Together with the statistical analysis, we then have a good argument for the claim to hold true. This would also be helpful for many derivations in statistical analysis where often uncorrelatedness is sufficient. For analysis, we need to look at the distribution of the true intermediate probabilities $P_{r,i}$, $i = 2, ..., m-1$, given the intermediate probability of the previous subset $p_{r,i-1}$. With many repetitive runs of SuS, enough necessary information will be available for this approach. Then we can plot the corresponding distributions considering all runs, calculate the empirical correlation of the samples or test for independence directly. This is not straightforward to do, since the true probabilities are not directly available but the object of interest. For this purpose, we create one result by SuS with $p_0 = 0.1$ and $N = 500,000$ and save all limit state values. Such a simulation yields quite accurate results and we can use the limit state values to assign threshold values to a probability, $F : \mathbb{R} \to [0, 1]$, $F(b) := p$. The corresponding function is the empirical cdf of $g(X)$ and allows to evaluate the approximately true intermediate probabilities in low N SuS runs by assigning the thresholds of the individual subsets an approximately true probability and dividing the probabilities of successive subsets

$$p_{r,i} = \frac{F(b_i)}{F(b_{i-1})} \ .$$

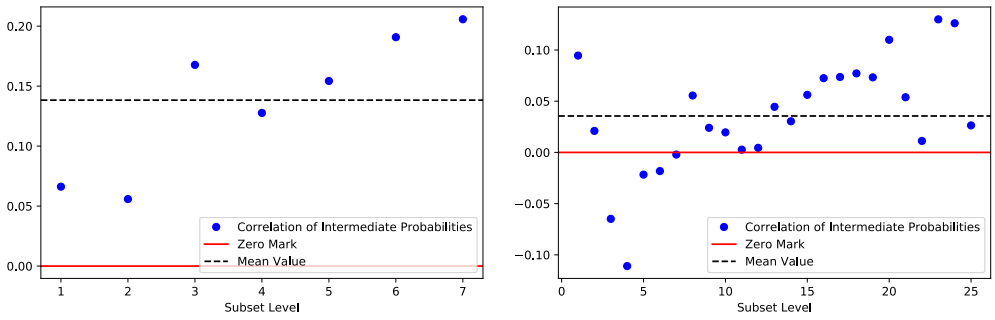
This way, we can derive an empirical distribution of the intermediate probabilities given the outcome of the previous subset $P(P_{r,i}|P_{r,i-1})$, $i = 2, ..., m-1$ and also directly evaluate the correlation coefficient of the two variables $\mathrm{Corr}(P_{r,i}, P_{r,i-1})$. This can be done with respect to a specific subset level or over all subsets. The last subset needs to be taken out of analysis, because its properties substantially differ. First, to prove independence of successive intermediate subset probabilities (Proposition 4.2.16), it was necessary to assume either independence of within subset samples which is not realistic under feasible computational efforts or that the dependency structure of samples does not significantly depend on the intermediate probability of the corresponding subset as described in Remark 4.2.14. Thus, we first need

Figure 4.10: Distribution of intermediate subset probabilities with respect to a previously realized intermediate probability $p_{r,3} \in [0.08, 0.1]$ or $p_{r,3} \in [0.1, 0.12]$ for subset level 4, based on 1000 independent simulation runs. Implementation (I4) was used with $p_0 = 0.1$ and $N = 500$.



to check whether such dependencies are present and we need to follow Remark 4.2.15 to destroy them. Thus, we start with an analysis based on implementation (I4). If not stated otherwise, results are with respect to limit state function g_1 (Equation 4.2). A first impression, including the whole distribution, is given in Figure 4.10 which at least shows no big differences, but also does not allow to draw certain conclusions. Additionally, a more abstract but also more insightful illustration is given in Figure 4.11, where we look at the empirical correlation coefficients with respect to the subset level. This time, $p_0 = 0.1$ and $p_0 = 0.5$ are considered. Figure 4.11 shows that for $p_0 = 0.1$ correlations are present, positive and appear to increase with higher subset levels. As there are correlations for $p_0 = 0.1$, Remark 4.2.14 can not hold in general. However, this well relates to the comments in Remark 4.2.14 and therefore is not a contradiction to the claim of independence as in Proposition 4.2.16 and Remark 4.2.17. The positive correlations can be explained by an increased bias due to decreased number of effective subset samples when sampling already starts with high dependent samples. Extreme cases will thus spread to higher subset levels (also compare Au and Wang (2014)). In case of $p_0 = 0.5$, the result is not as clear. Therefore, we here proceed with 5000 independent repetitive SuS runs and derive 95% confidence in-

Figure 4.11: Correlation coefficients with respect to the subset level, based on 1000 independent simulation runs and implementation (I4*).

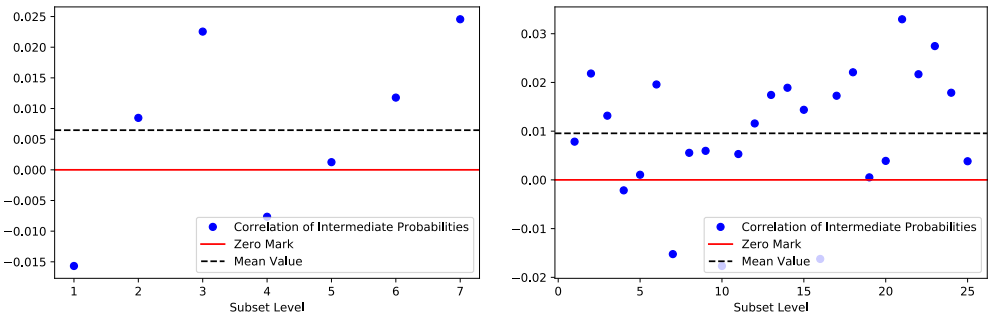


(a) $p_0 = 0.1$, $N = 500$ (b) $p_0 = 0.5$, $N = 300$

tervals for the correlation coefficients, resulting in e.g. $\rho_{11} \in [-0.035, 0.020]$ for subset level 11 and $\rho_{21} \in [0.043, 0.098]$ in level 21. So, in both cases we could not show the intermediate subset probabilities to be uncorrelated although $p_0 = 0.5$ seems to produce less correlated intermediate probabilities. As already stated, this can result from dependencies between seed samples affecting the intermediate subset estimation. This also explains why the correlations are positive since higher intermediate probabilities typically appear for higher dependencies because then the bias would increase, starting with more dependent samples in the next subset again, thereby inducing a higher expected probability. Extreme cases will thus spread to higher subset levels (also compare Au and Wang (2014)). Additionally, we did not drop the threshold seeds which might lead to a discrepancy between theory and simulation and did not apply the bias correction, which could also be done locally in every subset individually.

Although the effects might be negligible for higher p_0, we can now also follow Remark 4.2.15 to hopefully get the desired results for high and also small p_0. To do so, the same experiment is performed, but this time only 10 seeds are kept and used as a start for MCMC. The other seeds are dropped and not used in the next subset. The idea is to destroy the dependency structure between those samples, randomly removing many of them until only 10 seeds are left, which are then used for MCMC. Additionally, we increase the number of samples per subset to $N = 1000$ so that a destruction of the dependency structure of the seeds is even more likely. Also, we

Figure 4.12: Correlation coefficients with respect to the subset level, based on 3000 independent simulations runs and implementation (I5*).

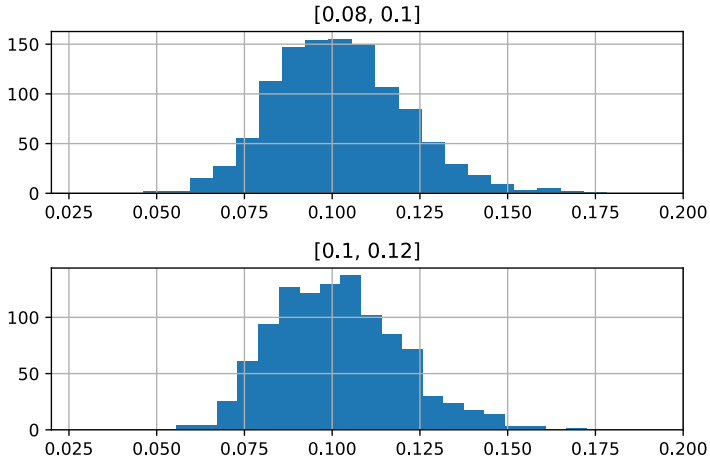


(a) $p_0 = 0.1$, $N = 1000$, $N_s = 10$ (b) $p_0 = 0.5$, $N = 300$, $N_s = 20$

drop threshold related sample seeds for better theoretical consistency, having only seeds in the stationary distribution. The result is shown in Figure 4.12 where intermediate subset probabilities appear to be uncorrelated now. This time, we also got evidence on uncorrelated intermediate subset probabilities by the corresponding confidence intervals. For $p_0 = 0.1$, all subset levels 1,3,5,7 include $\rho = 0$ such as $\rho_1 \in [-0.051, 0.020]$, $\rho_3 \in [-0.013, 0.058]$, $\rho_5 \in [-0.035, 0.037]$ and $\rho_7 \in [-0.011, 0.060]$. Empirical distributions (compare Figure 4.13) support the assumption of independence. In general, small dependencies might still be present due to keeping $10 > 1$ samples, we however might assume that these can be neglected if the dependency structure of the seeds is destroyed appropriately. Our aim also is to show that not all seeds except for one need to be dropped to receive the desired result, allowing an outlook for application in practice, not obviously causing too bad ergodicity issues in comparison to having only one seed sample remaining (compare Remark 3.1.14). For $p_0 = 0.5$ we did not increase the sample number and kept $N_s = 20$ seeds, as results seemed to almost provide evidence for uncorrelated intermediate probabilities, without modification, already. We still might have a small correlation present, but $\rho = 0$ is included in the corresponding confidence intervals for almost all subset levels. Thus, to guarantee the often, particularly in statistical analysis, beneficial property of having approximately uncorrelated (most likely even independent) intermediate probabilities, following Remark 4.2.15 should meet the requirements.

Next, Table 4.5 demonstrates the effect of dropping threshold related

Figure 4.13: Distribution of intermediate subset probabilities with respect to a previously realized intermediate probability $p_{r,3} \in [0.08, 0.1]$ or $p_{r,3} \in [0.1, 0.12]$ for subset level 4, based on 3000 independent simulation runs. Implementation (I5) was used with $p_0 = 0.1$ and $N = 1000$.



samples according to Remark 4.2.7 instead of keeping them as in current implementations. Additionally to dropping all samples with limit state value equal to the threshold, we also did not use samples with distances less than 3 chain members in their corresponding Markov chain. We see that keeping the threshold increases the bias and we have two effects that possibly cancel each other out in the ordinary SuS implementation, the positive bias by order statistics and the negative bias by keeping the threshold seeds. In total, we could conclude that it is thus better to keep the threshold samples anyhow, but it is not clear if the negative bias can also become bigger than the positive one. Then we had a systematic underestimation of the failure probability and not much information on how big it could become. Also, there still remains a positive bias, which especially for small N and high p_0 will be non-negligible. When several repetitive runs are performed and the bias remains the same but the CV is decreased, it can indeed play a significant role. Also note that these results were produced with good adaptive sampling, which may not be reachable in reality always. So a much higher bias is imaginable, depending on the problem at hand. An alternative, which possibly adds some control to the situation, is given in Table 4.6. Note, if thresholds are kept as in current implementations, the bias correction formula will not suit in the way

$\mathbf{p_0}$, **N**	$\mathbf{p_f}$	$\mathbf{1 \cdot 10^{-3}}$	$\mathbf{5 \cdot 10^{-5}}$	$\mathbf{1 \cdot 10^{-6}}$	$\mathbf{5 \cdot 10^{-8}}$
$0.1, 500$	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, keep	1.04(4)	5.39(8)	1.11(11)	5.78(16)
	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, drop	1.07(7)	5.57(11)	1.18(18)	6.27(25)
	CV **(keep)**	0.34	0.47	0.62	0.75
	CV **(drop)**	0.34	0.46	0.61	0.75
$0.5, 300$	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, keep	1.06(6)	5.35(7)	1.12(12)	5.65(13)
	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, drop	1.14(14)	6.13(23)	1.34(34)	7.29(46)
	CV **(keep)**	0.36	0.45	0.58	0.67
	CV **(drop)**	0.35	0.45	0.56	0.65

Table 4.5: **Subset Simulation with and without dropping threshold related seed samples**: We examine the relative bias (with respect to the true failure probability p_f) in percent ($\%B$) and the corresponding coefficient of variation, under varied true failure probability. The number of total evaluations is similar for both considered p_0. We see that the bias increases when dropping the threshold. The effect becomes stronger for higher p_0 and lower p_f. Coefficients of variation remain the same. Table values are based on $10,000$ independent simulation runs and linear limit state function g_1 (Equation 4.2). g_{2a} and g_{2b} (Equation 4.3-4.4) lead to similar results.

we apply it here because assumptions in theory are not fulfilled then. So both alterations of the algorithms go together best, which we show next.

$\mathbf{p_0}$, **N**	$\mathbf{p_f}$	$\mathbf{1 \cdot 10^{-3}}$	$\mathbf{5 \cdot 10^{-5}}$	$\mathbf{1 \cdot 10^{-6}}$	$\mathbf{5 \cdot 10^{-8}}$
$0.1, 500$	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, drop	1.07(7)	5.57(11)	1.18(18)	6.27(25)
	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, corr.	0.99(−1)	4.83(−3)	0.98(−2)	4.9(−2)
	CV*	0.36	0.51	0.72	0.94
	CV*, **corr.**	0.34	0.45	0.60	0.74
$0.5, 300$	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, drop	1.14(14)	6.13(23)	1.34(34)	7.29(46)
	$\hat{\mathbf{p}}_\mathbf{f}(\%B)$, corr.	0.98(−2)	4.86(−3)	0.96(−4)	4.82(−4)
	CV*	0.40	0.55	0.75	0.95
	CV*, **corr**	0.34	0.44	0.54	0.62

Table 4.6: **Subset Simulation with and without bias correction**: We examine the relative bias (with respect to the true failure probability p_f) in percent ($\%B$) and adjusted coefficient of variation CV*. CV* yields the CV with respect to true p_f and therefore accounts for the increased, or for the bias corrected result decreased, variance as a consequence of a different reference value under constant coefficient of variation. Table values are based on $10,000$ independent simulation runs and linear limit state function g_1 (Equation 4.2). g_{2a} and g_{2b} (Equation 4.3-4.4) lead to same results. The number of total evaluations is similar for both considered p_0.

The effect by bias correction can be significant, in particular for high p_0, low N as well as low p_f. Under joint bias correction and dropping of the threshold, we achieve less biased results in comparison to the original method, in all simulations. For higher N the results remain the same, just on a lower and less important scale. Remember also that the slightly negative bias was expected, because we neglected the change in distribution according to Remark 4.2.13 in our formula derivation. However, despite having achieved good results with the empirical approach it is not absolutely clear if this is the right way to handle the dependent order statistics (compare Remark 4.2.30). By comparing the results for dropping the threshold samples and bias correction with Table 4.5, it seems that the current implementation and the new one with bias correction do not differ so much in the result. However we might have reliability problems that lead to worse adaptive sam-

pling and thus a higher bias due to the resulting less effective samples. Then the bias might become even higher when keeping the same sample number per subset N, making an understanding of the bias even more important and valuable. Most importantly, it might provide very accurate results by many independent runs of the algorithm, if necessary, and does help to understand the fundamental properties of the algorithm.

4.2.4 Discussion

Note, in contrast to analysis by order statistics, a general Bayes approach with updates by Binomial distributed information will not deliver an exact result, in particular for small N.

Remark 4.2.31 (General Bayesian Approach: Questionable Threshold Selection)**.** *If we aim to derive distributions of intermediate probabilities $p_{r,i}, i = 1, ..., m-1$ by a general Bayesian approach, starting with a non-informative prior such as* $\mathrm{Beta}(0,0)$ *(most uncertainty and no zeros after estimation guaranteed by subset construction) and identifying information in each subset by the given failures $p_0 N$, then the posterior distribution is a Beta distribution* $\mathrm{Beta}(p_0 N, (1-p_0)N)$*. However, although we may derive the probability distribution accurately, we do not know which threshold b_i it has to be assigned to. The Beta distribution corresponds to a threshold b_i^* where in each subset a Binomial experiment is performed. As a result of this experiment, we should have $p_0 N$ limit state values below the threshold b_i^* and $(1-p_0)N$ above it. But, by definition this case never happens in reality because we always hit the threshold, by defining it according to the limit state value of a sample. A suitable threshold b_i^* on the other hand has to separate limit state values of samples in the subset and the ones out of it. Thus it has to be placed somewhere in between $g^{r,i}_{(p_0 N)}$ and $g^{r,i}_{(p_0 N+1)}$, where exactly is unclear. If doubles occur the situation becomes even more complicated.*

Regarding the findings of this section, in particular the consequences of independent intermediate subset probabilities are far-reaching, allowing to proceed with theoretical analysis under a strong foundation and to better judge effects of altered parameters in SuS.

Remark 4.2.32 (Comparison to Previous Analysis)**.** *We figured out that the bias is due to the construction of the subsets with order statistics. In*

past analysis the bias was considered a consequence of correlated intermediate subset probabilities mostly. However, having shown that those should be uncorrelated, the latter explanation may not be suitable. The same holds for upper and lower boundaries on the variance. The increased variance with respect to higher p_0 which had to be explained, is more likely the consequence of a decreased effective sample number N (increased $\bar{\gamma}$) in evaluation for higher p_0 as discussed in Section 4.1.3.

Remark 4.2.33 (Additional Bias by Stopping Criterion)**.** *By definition of Subset Simulation, we may hit the final threshold b^* much earlier than we do on average, possibly even in the first subset. Once we have $p_0 N$ samples with limit state value smaller than b^*, we directly terminate the algorithm. So, we had $m-1$ Monte Carlo simulations that could have end the simulation before, when finally getting to m. This bias, produced in every subset $\mathbb{F}_i$, $i < m$, is given by*

$$\sum_{k=p_0 N}^{N} \frac{k}{N} p_0^i \text{binom}\left(k; N, p_f / \prod_{j=1}^{i-1} p_{r,j}\right)$$

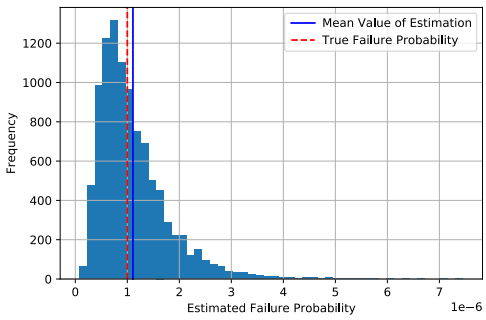
with $\text{binom}\left(k; N, p_f / \prod_{j=1}^{i-1} p_{r,j}\right)$ *being the probability mass function of the binomial distribution for k successes in N independent Bernoulli trials with probability $p_f / \prod_{j=1}^{i-1} p_{r,j}$. The fact that k starts at $p_0 N$ only allows estimated probabilities greater than $p_0 N$ so that a termination of the algorithm by the stopping criterion is reached, $\frac{k}{N} p_0^i$ the estimated value and*

$$\text{binom}\left(k; N, p_f / \prod_{j=1}^{i-1} p_{r,j}\right)$$

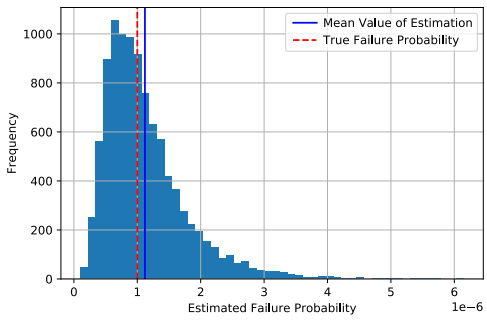
the corresponding probability of the event $\{g^i_{(k)} \leq 0\}$. Unfortunately, we neither know the true failure probability p_f, nor the true intermediate subset probabilities $p_{r,i}$, $i = 1, ..., m-1$, when applying Subset Simulation for reliability evaluation. Thus, we can, in addition to the uncertainty from an unknown effective sample number per subset N, only approximately cancel this bias. Nevertheless, having the opportunity to derive a distribution for P_f after termination of Subset Simulation by Corollary 4.2.20, we can approximate the total bias from the stopping criterion by

$$\sum_{i=1}^{m-1} \sum_{k=p_0 N}^{N} \frac{k}{N} p_0^i \int_{0<p\leq 1} \text{binom}(k; N, p) \cdot \text{Beta}(p; a, b) dp \ ,$$

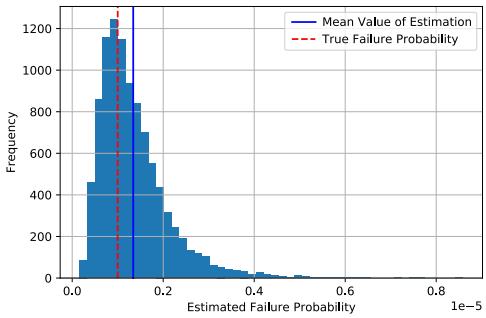
Figure 4.14: Shape of the Distribution of SuS Results. For derivation of the results, 10,000 independent simulations of SuS were performed, respectively. Here, we used the implementations (I4*), (I5*) and (I6*).



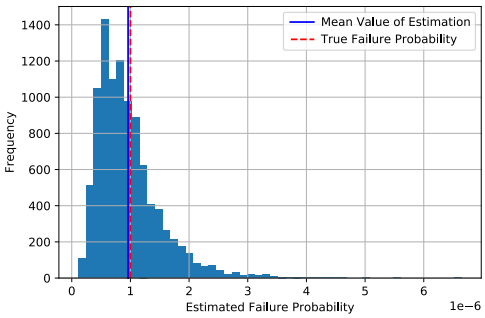
(a) $N = 500$, $p_0 = 0.1$, (I4*)



(b) $N = 300$, $p_0 = 0.5$, (I4*)



(c) $N = 300$, $p_0 = 0.5$, (I5*)

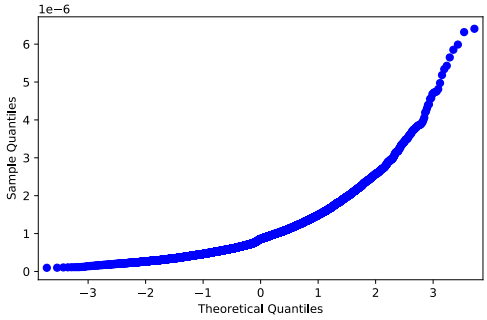


(d) $N = 300$, $p_0 = 0.5$, (I6*)

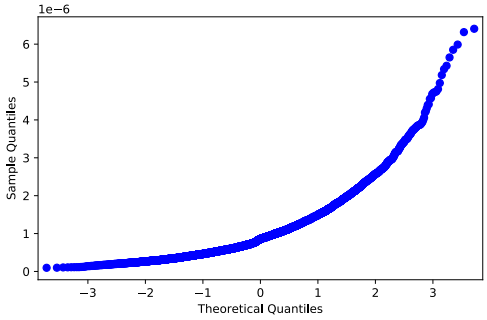
replacing the unknown probabilities by the best estimates available. The parameters of the Beta distribution are given by Corollary 4.2.20.

Lastly, we would like to add some information on the distribution of SuS results, to get a more holistic understanding of SuS results with small sample numbers. Illustrations are given in Figure 4.14. The distributions appear to be non-normal where a qq-plot of the data allows more precise classification of them as shown in Figure 4.15. Figure 4.15c and Figure 4.15d indicate that the distribution of SuS results matches a lognormal distribution well for small N, most likely even independent of p_0. The first part of this conclusion coincides with Breitung (2019).

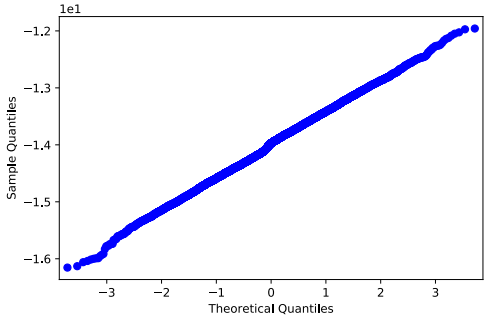
Figure 4.15: qq-plots of SuS results, based on $10,000$ independent simulations of SuS with (I4*).



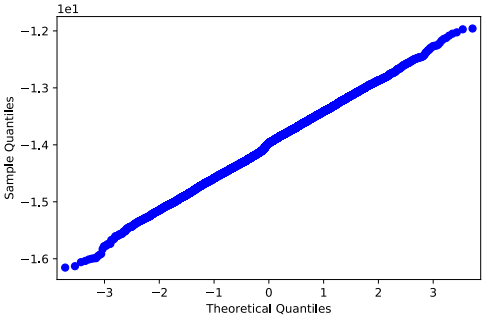
(a) qq-plot of $\hat{p}_f$, $p_0 = 0.1$, $N = 500$



(b) qq-plot of $\hat{p}_f$, $p_0 = 0.5$, $N = 300$



(c) qq-plot of $\log(\hat{p}_f)$, $p_0 = 0.1$, $N = 500$



(d) qq-plot of $\log(\hat{p}_f)$, $p_0 = 0.5$, $N = 300$

4.3 Conclusion: Practical Enhancements

Our theoretical analysis and simulation studies have provided insights into fundamental properties of Subset Simulation, allowing for further understanding and explicit goal-oriented modifications. In short, we have made the following discoveries.

1. High p_0 does not necessarily result in higher variation of the result if sample seeds are re-used in successive subsets. Thus, opportunities for extension or modification of the algorithm are not limited to low p_0 values, offering new fields of relevant study. In particular our novel algorithm, presented in Chapter 5 will benefit a lot from this finding.

2. Simulations have shown that the chain length plays a major role in efficiency, especially for high choices of p_0. This partially explains the higher than expected coefficients of variation for high p_0 in the past.

3. We derived a transformation for calculating an unbiased Subset Simulation result which works under minor conditions, instead of an $O(1/N)$ unbiased one that is derived under asymptotic arguments. Limits of applicability are explicitly given.

4. The typical threshold selection in Subset Simulation was found to lead to a negative bias. Here, further studies should examine if the resulting SuS bias can also become negative. If it can, those cases must be identified to carry out secure and reliable analysis of failure probabilities under small N when threshold samples are kept as in most current implementations.

5. Successive intermediate subset probabilities have been theoretically shown to be independent under weak assumptions. Practically, high p_0 may give uncorrelated intermediate subset probabilities and if the implementation is slightly modified both, low and high p_0, should lead to approximately uncorrelated intermediate probabilities. This has crucial consequences for a more closed statistical analysis and suitable findings by simulation.

6. Effective sample numbers, given sample numbers per subset N, were demonstrated to highly depend on the choice of p_0. Higher p_0 result

in a lower number of effective samples. This, together with the automatically reduced chain length explains an inconsistency of theory and findings in simulations with respect to the coefficient of variation being higher than expected for higher choices of p_0.

Subset Simulation: Guidelines for Application

The following points shortly summarize the parameter settings we identified as most important ones for good efficiency during our simulation studies:

- It is important to use enough chains to have a higher probability to find all failure regions. A specific number can not be given here.
- Implementations with adaptive sampling (I3,I4,I5,I6) are much more efficient than non-adaptive ones (I1,I2) and should preferably be chosen.
- The proposal spread should not be updated within Markov chains. Otherwise an additional bias is the consequence.
- The distribution of SuS results highly depends on the sample number. In particular, when deriving confidence intervals or when interpreting the rRMSE, care has to be taken.
- It is advisable to use several independent runs of SuS for one reliability evaluation because theory can often not provide secure results (in particular for the distribution) in advance and evidence through multiple trials is therefore very valuable.
- The chain length should be set to a fixed value such as N_l between 10 and 20, independent of p_0.
- Intermediate subset probabilities $p_0 \geq 0.1$ do not result in a lower efficiency necessarily and can be used as well. A recommendation on values $p_0 \geq 0.8$ is not given by simulation evidence, since higher p_0 were not included in our practical tests.
- Keeping threshold samples seems to be okay and even beneficial, if no bias correction is applied. This however needs further analysis, to avoid an unwanted negative bias.

Note that the last three points have not been identified so far, where the other points have already been discussed in e.g. Au and Beck (2001b), Au and Wang (2014), Papaioannou et al. (2015), Breitung (2019).

Chapter 5

Subset Simulation Interpolation: Introduction and Statistical Analysis[1]

In Chapter 3 we have seen that SuS yields the failure probability for one fixed constellation of the basic variables where analysis of its statistical properties has been extended in Chapter 4. To evaluate the failure probability when the distribution of a random variable changes, a possibly inaccurate approximation by Taylor expansion or an additional calculation is necessary. A short overview of existing Subset Simulation approaches with the potential for efficient evaluation of dynamic models and a clarification on the difference to the method of this section was given in Chapter 3. However, if many different distributions of a specific random variable are relevant, then either the approximations might become unacceptable or repetitive calculations can result in very high, possibly infeasible, computational demands. A main drawback of local approximations is that their accuracy is generally unknown, especially if the different distributions differ substantially from each other. In this chapter, we propose a novel algorithm, 'Subset Simulation Interpolation' (SuSI), which is based on Subset Simulation and deals with the solution of this problem.

In particular, we describe the idea as well as the implementation of SuSI and give a detailed statistical analysis in this chapter. A simulation study

[1]The algorithm was first presented at ESREL2019 (Blandfort et al. (2019a)). Since then it has been further developed, applications were extended and a theoretical foundation was given. The result is presented in this dissertation.

and its extensive application opportunities are presented in Chapter 6. The results of both chapters taken together (as suggested by Remark 2.1.3), give a good indication of SuSI yielding a similar computational demand as a static SuS evaluation, in many cases.

5.1 Introduction of the Algorithm

In this section, we explain how the new algorithm deals with the problematic computational demands given in reliability evaluation of dynamic models as described in Section 2.2. We start with a brief introduction of the underlying idea, allowing to understand how the approach can provide more information about structural failure than ordinary SuS. In the next sections, we add a detailed description, explicitly state its implementation and derive general statistical properties. Furthermore, theoretically derived results and the given concept are tested by practical simulations, verifying efficiency and functionality of the algorithm.

5.1.1 A Method Designed for Dynamic Models

Subset Simulation Interpolation is designed for dynamic models such as introduced in Section 2.2. To better trace efficiency, we next make the dynamic model representation more explicit. Assume we have n relevant distributions of the dynamic variable $\mathbf{X}_k$ and have to examine the corresponding failure probability of the structure as in Equation 2.2 by

$$p_f = \int_{D_k} P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y) f_k(y) dy \tag{5.1}$$

for every pdf $f_k \in \{f_k^1, ..., f_k^n\}$. The two assumptions, which have to be fulfilled by the corresponding static reliability problem, are independence of the dynamic variable $\mathbf{X}_k$ with respect to all other random variables (compare Assumption 2.2.1) and monotonicity of the limit state function with respect to it (Assumption 2.2.2). Our proposed algorithm SuSI was developed to deal with exactly this type of dynamic model and yields the following benefit.

Remark 5.1.1 (Benefit One of SuSI)**.** *If we are given a limit state function g and stochastic properties $\mathbf{X}$ as in Section 2.2 and want to evaluate the probability of failure for n different distributions of $\mathbf{X}_k$, SuSI allows to derive the result of all those failure probabilities by one single simulation run. In contrast, SuS needs n repetitive simulation runs or utilization of local approximations with unknown accuracy.*

Remark 5.1.2 (Benefit Two of SuSI)**.** *Another benefit is the opportunity to compute a single SuS result by different underlying limit state functions so that more robust SuS results might be determined (cf. Section 6.2.3)*

The comparison of SuSI and SuS with respect to the computational demand comes natural because both perform best at the same type of reliability problems where the proposed method is specialized on dynamic models. Remark 5.1.1 and Remark 5.1.2 clarify that SuSI does not necessarily need to beat the performance of SuS in general to improve efficiency in the dynamic model setting. Depending on how much evaluations with different distributions of $\mathbf{X}_k$ are required, it would even improve efficiency if the performance was much worse in the static setting. In short, we, although not entirely accurate, conclude:

Computational demands of SuSI can be n times lower than that of SuS if both have the same efficiency in a single static reliability evaluation.

This is a good guideline, but one should also remind that it neglects some important facts such as the possibility to consider approximative methods for dynamic models. On the other hand, such methods might not deliver results as accurate as SuSI does. At least, we will be able to interpret the computational demand easily by deriving the cost in relation to an ordinary static SuS evaluation, which is exactly the type of information we need to rate the efficiency of SuSI. The efficiency of the algorithm will be studied theoretically in Section 5.3 and practically in Section 6.1 where we will always refer to the above statement in so far that efficiency is evaluated with respect to the efficiency of a static SuS evaluation.

5.1.2 Idea

In a nutshell, SuSI estimates a failure probability by a pdf weighted sum (or integral) of SuS estimates according to Equation 5.1 as illustrated in Figure 5.1. These single SuS estimates are not independent, instead they are created by one SuS-like simulation run and are given by the product of the intermediate subset probabilities up to one specific subset.

This might seem to be a minor change at first, but the intermediate subsets are created by variation of the limit state function differently than in SuS. By altering the way subsets are adaptively defined, every of those SuS estimates yields a conditional failure probability with respect to a fixed value of the dynamic variable $\mathbf{X}_k$. Still, this different creation of subsets uses the same basic idea as in SuS and keeps most remarkable properties for reliability evaluation. So, under an increased gain of information by the algorithm, the

principles of SuS are kept as a basis.

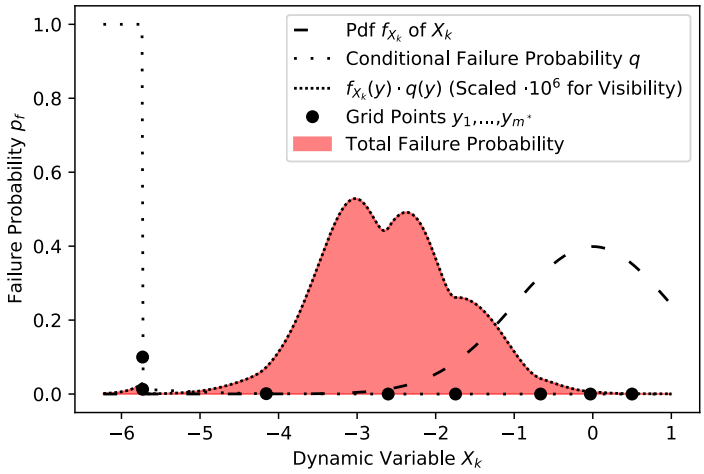
Remark 5.1.3. *The similarity between SuS and SuSI is highly beneficial for the new algorithm, as most enhancements for Subset Simulation are also applicable in our algorithm. As a popular method, it has been enhanced and tested a lot over the past two decades, yielding a strong foundation for analysis and proof of concept for SuSI also.*

In more detail, the idea and substantial difference of SuSI with respect to SuS is twofold:

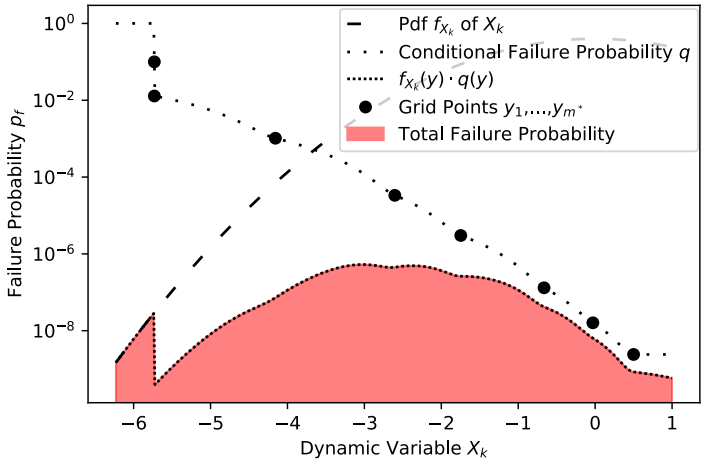
(W1) The first difference is the way subsets are created: In SuSI, the threshold for subset generation is kept constantly set to failure $b = b^*$ and the values of a specific random variable $\mathbf{X}_k$ are changed instead. Doing so, we get the conditional failure probabilities $P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y)$ with respect to (deterministic) values $y \in D_k$ of $\mathbf{X}_k$. If monotonicity (Assumption 2.2.2) is fulfilled, this allows us to construct a sequence of intermediate events $\mathbb{F}_1 \supseteq \mathbb{F}_2 \supseteq ... \supseteq \mathbb{F}_{m^*}, m^* \in \mathbb{N}$ as in SuS, where in contrast to SuS, we define $\mathbb{F}_i := \{\mathbf{x} \in D | g(\mathbf{x}) < b^* \wedge \mathbf{x}_k = y_i\}$ with $y_1 < y_2 < ... < y_{m^*}$ and $y_1, y_2, ..., y_{m^*} \in D_k$ for monotone increasing g with respect to $\mathbf{x}_k$ and refer to subsets by $\mathbb{F}_{y_i}, i = 1, ..., m^*$ instead of $\mathbb{F}_i$ to clarify the dependency on the value of the dynamic variable. If a subset is either produced by SuS or SuSI, we keep the original notation and refer to it by $\mathbb{F}_i$. Therefore, in SuSI, only failure probabilities ($b = b^*$), and not the probability of non-critical threshold violations, are calculated at every subset. The number of subset levels might also differ from that given in SuS and is referred to as $m^* \in \mathbb{N}$. An illustration of the subset structure and the corresponding failure space is given in Figure 5.2. Figure 5.3 provides a comparison with respect to SuS. Additionally, because this is the most important point of the idea, Figure 5.4 again illustrates the difference in subset creation of SuS and SuSI.

(W2) Second, the discrete conditional failure probabilities, given at each grid point, are extended by a subsequent interpolation. This creates a continuous conditional failure probability estimation function $q : D_k \to [0, 1]$, $q(y) = P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y)$ with respect to the dynamic variable, as shown in Figure 5.5. After interpolation, one can then evaluate (5.1) by a simple one-dimensional integration. The main

Figure 5.1: Subset Simulation Interpolation: Estimating the Failure Probability by weighted SuS estimates. The result of SuSI is the result of an integration of the function $f_k(y) \cdot q(y)$ over the domain of the dynamic variable $\mathbf{X}_k$ (compare Equation 5.1). Correspondingly, we have the conditional failure probability function q which is constructed by SuS like estimates weighted with the pdf of $\mathbf{X}_k$. For demonstration, we have chosen g_1 (compare Equation 4.2) with 4 random variables as a limit state and assumed $p_f = 1 \cdot 10^{-6}$ for the failure probability. To stick to the typical monotonicity assumption (Assumption 2.2.2), we dropped the negative sign before the Gaussian variables in g_1. The result is shown for linear and for log scale because of q decreasing exponentially.



(a) Linear Scale (And Scaled $f_k(y) \cdot q(y)$)



(b) Log Scale

benefit is that this inexpensive evaluation can then be repeated for several different distributions of $\mathbf{X}_k$ instead of the typically extremely computational demanding reliability evaluation.

Note that the failure sets in Figure 5.2b are 2-dimensional, thus are of a lower dimension than in SuS (compare Figure 5.3). In terms of exploring the failure space, we well see that SuSI tries to sample along the failure space with respect to $\mathbf{X}_k$, exploring information about the failure space in each subset.

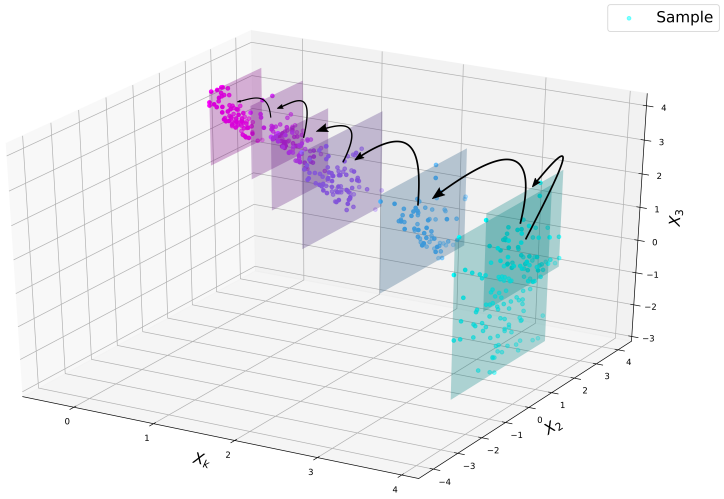
For clarification of the close connection between SuS and SuSI, we can even interpret SuS as a special case of SuSI.

Remark 5.1.4 (SuS as Special Case of SuSI)**.** *If we look at the general formulation of SuS, variation of the threshold at each subset level $b_1 > b_2 > ... > b_m = b^*$ can also be understood as variation of the limit state according to $g_i(\mathbf{X}) := g(\mathbf{X}) - b_i$ with $g(\mathbf{X})$ the original limit state function. In comparison to that, SuSI would interpret the exact same constellation as $g_i(\mathbf{X}) := g(\mathbf{X}) - \mathbf{X}_{d+1}$, where $\mathbf{X}_{d+1}$ is then understood as a new artificial variable set to a different deterministic value in each subset adaptively. $\mathbf{X}_{d+1}$ takes the role of the dynamic variable $\mathbf{X}_k$. The probability of failure is then given for a constant $\mathbf{X}_{d+1} = b^*$, so that the conditional failure probability at this value is just taken as estimation of the failure probability, not needing to calculate a linear combination of several estimates to get the result. In terms of the pdf f_k, this would refer to $\mathbf{X}_k = b^*$ almost surely, which can be approximated by a Dirac delta function centered at b^*. In this setting, we will not benefit from SuSI, just introducing the potential bias for interpolation if we do not directly evaluate conditional on $\mathbf{X}_{d+1} = b^*$. Also the general approach for finding the deterministic values of $\mathbf{X}_{d+1}$ for each subset will yield additional unnecessary computational demands if we follow the standard procedure. Additionally, no double evaluations, as discussed later, are necessary. More details follow in Section 5.2. So SuS is a special case of SuSI that utilizes computational benefits of a specific setting.*

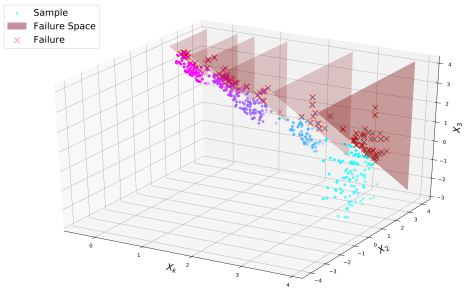
The basis of both algorithms is an adaptive creation of subsets, but the resulting effect by their differences in creation is not a priori clear. The following remark comments on some consequences.

Remark 5.1.5 (Subset Creation: Comparison of SuS and SuSI)**.** *The most important difference comes from the fact that in SuSI, the failure sets change*

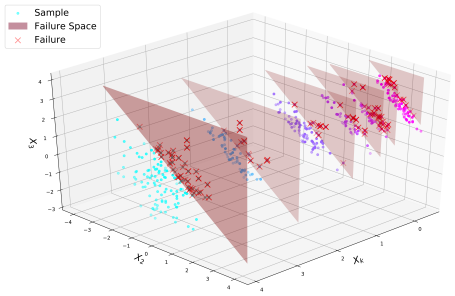
Figure 5.2: Subset Simulation Interpolation: Dimension Reduction and Exploring the Failure Space. SuSI changes $\mathbf{X}_k$ in each step so that we produce samples that belong to points on the planes shown in Figure 5.2a. The planes were chosen so that they contain all sample values of a subset. In Figure 5.2b and Figure 5.2c the corresponding failure regions are drawn and failed samples are marked red. This well shows the gain of information by SuSI in every subset, receiving samples which belong to the failure region always. The limit state function chosen for illustration is g_1 (compare Equation 4.2, monotone decreasing with respect to component $\mathbf{x}_k$) with 3 random variables and $p_f = 1 \cdot 10^{-3}$, having a starting subset corresponding to $\mathbf{X}_1 = \mathbf{X}_k = 3.9$. Note the double plane at $\mathbf{X}_k = 3.9$ corresponds to starting the simulation with SuS until enough failures are found (compare Section 5.2).



(a) Subset Construction



(b) Failure Sets, Perspective 1



(c) Failure Sets, Perspective 2

Figure 5.3: Subset Simulation: Exploring the Failure Space. The limit state function is the same as in Figure 5.2, g_1 (compare Equation 4.2) with 3 random variables and $p_f = 1 \cdot 10^{-3}$. In contrast to SuSI, the failure set is 3-dimensional and does not change in successive subsets. Moreover, $\mathbf{X}_k = \mathbf{X}_1$ does not take a special role. Starting around $(0, 0, 0)$ with highest probability density, we sample in the direction of steepest descent by MCMC to find the failure region.

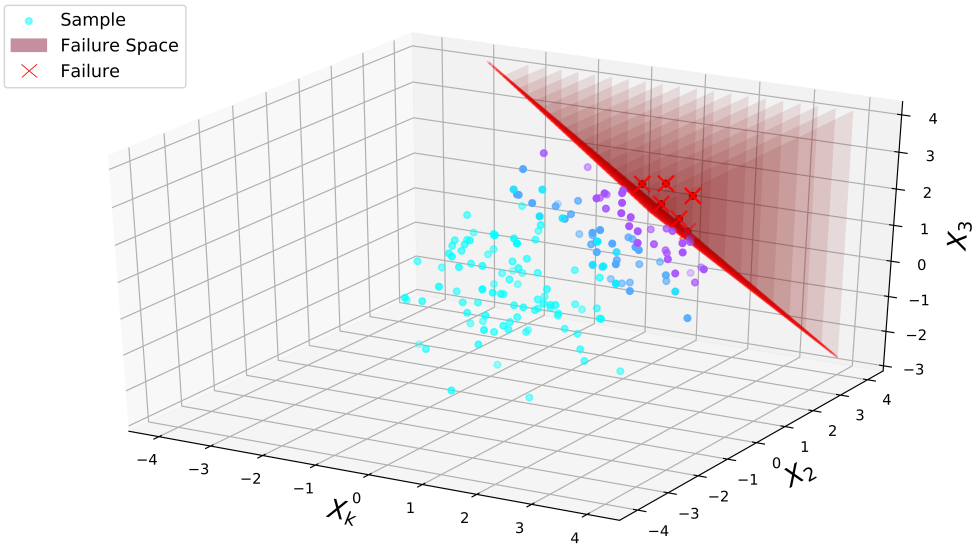
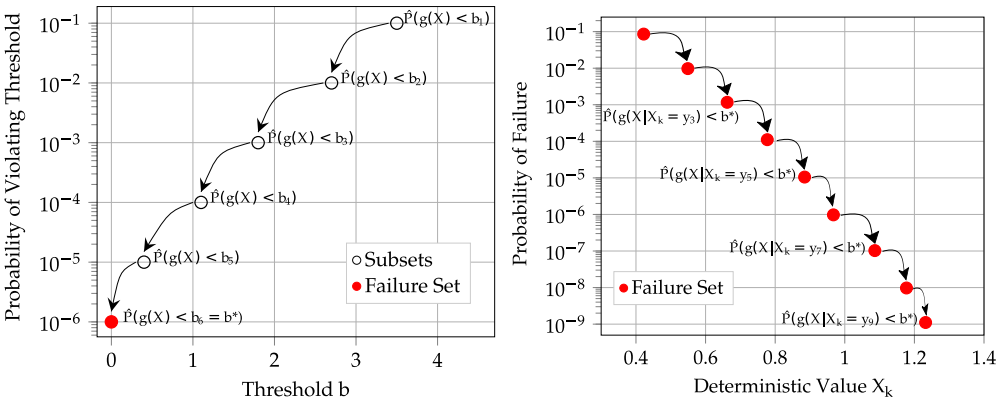
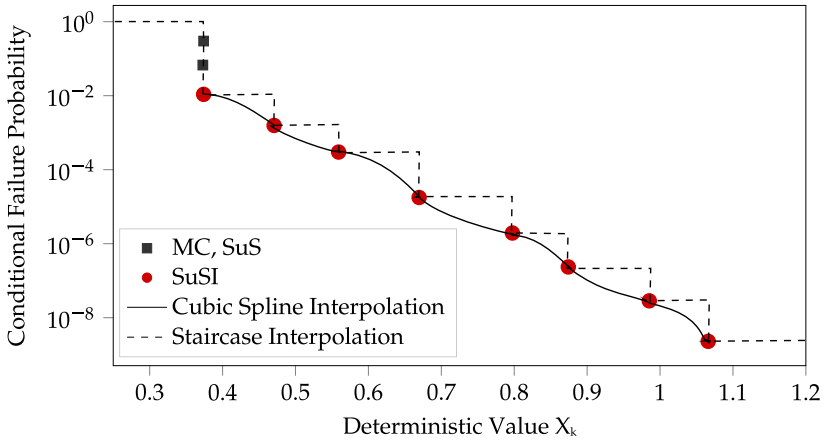


Figure 5.4: Comparison of the subset generation procedure in SuS and SuSI. In contrast to SuS, where subsets are defined with respect to successive thresholds, SuSI conditions on deterministic values of the dynamic variable $\mathbf{X}_k$. This way, SuSI explores information about failure conditional on $\mathbf{X}_k$ in every, or many, subsets, depending on the importance of $\mathbf{X}_k$ on failure.



(a) SuS (b) SuSI

Figure 5.5: Blandfort et al. (2019a), modified. SuSI with upper staircase and cubic spline interpolation: Extension of the discrete set of evaluated conditional failure probabilities in SuSI to a continuous result.



in each subset whereas in SuS the conditional failure sets change. This results in a substantial difference of application of the algorithm and also makes the main idea to provide more information by a single run of the algorithm because this way failure sets are examined in each subset. The properties of both methods, should however coincide in many cases. With respect to ergodicity, we implicitly assume at start of the algorithm that the direction of important regions remains the same for higher subset levels. The equivalent in SuS is given by a change of the conditional failure sets in each subset by variation of the threshold b. Thus the underlying assumptions for receiving correct results are similar, both assuming that the lower level subsets allow to extrapolate to higher level subset failure sets as discussed in Breitung (2018) and Breitung (2019). As a special feature of SuSI, the dimension of the state space is reduced by one, in comparison to ordinary SuS. This does not necessarily affect efficiency, but it can, if only higher failure probabilities are given more weight or if it alters the limit state function in a beneficial way. Note, this also potentially decreases efficiency.

In order to compare the efficiency of SuS and SuSI next, we need to analyze the total number of evaluations E_T and therefore study an implementation of SuSI. Thus, we next discuss how SuSI can be implemented efficiently and emphasize differences between SuS and SuSI.

5.2 Implementation

Regarding the twofold idea (compare (W1) and (W2) in Section 5.1) of SuSI, we next state challenges that come with such an approach and explain how to modify the procedures applied in SuS to efficiently deal with them, providing Algorithm 2[2]. The challenges can be directly formulated with respect to the twofold idea given above:

(C1): Challenges in subset creation (W1).

- (C1-1) Selection of y_1. It is not a priori clear how to select the first value y_1 of the dynamic variable to define the first subset. Depending on the limit state function, it might even be impossible to select a $y_1 \in D_k$ so that the initial conditional failure probability $P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y_1)$ is high enough for efficient evaluation by Monte Carlo simulation.
- (C1-2) Selection of y_{i+1}. The subsets are successively defined by $\mathbb{F}_{y_i} := \{\mathbf{x} \in D | g(\mathbf{x}) < b^* \wedge \mathbf{x}_k = y_i\}$ with ascending $y_1 < y_2 < ... < y_{m^*}$. Efficient calculation as in Subset Simulation is typically achieved if the intermediate subset probabilities approximately take a specific pre-defined value p_0, $P(\mathbb{F}_{y_{i+1}} | \mathbb{F}_{y_i}) \approx p_0, i = 1, ..., m^* - 1$, or at least are in a specific range of probabilities such as $[p_l, p_u]$ with $p_l < p_u$, $p_l, p_u \in (0, 1)$. The corresponding y_i that provide these probabilities however are unknown in advance.

(C2): Challenges in interpolation and evaluation of the failure probability (W2).

- (C2-1) Interpolation method. Because the shape of the function q is generally unknown, one has to account for possible systematic overestimation or underestimation of the failure probability by interpolation. In particular, also a critical negative bias is generally possible.
- (C2-2) Coverage of domain. Regarding the dynamic variable $\mathbf{X}_k$, we choose a minimum value y_1 and a maximum value y_{m^*}, where the

[2] An implementation of the algorithm is publicly available at GitHub (https://github.com/FBlandfort/Subset-Simulation-Interpolation).

conditional failure probability is evaluated. In the desired evaluation of Equation 5.1 we can thus not cover the whole domain of $\mathbf{X}_k$ by the given evaluated points and need to extrapolate beyond the given minimum and maximum or neglect the corresponding $\mathbf{X}_k$ values.

On the whole, it might seem not worth with the effort to deal with these challenges, but we will show shortly how they can be overcome very well and how it is indeed worth the effort to do so. First, we need to allow for approximations, introducing a thereby permitted error.

Definition 5.2.1 (Approximation Accuracy)**.** *The maximal admissible error by approximation is given by* $\epsilon_{\max}$.

The approximation error $\epsilon_{\max}$ can be set to any value, where one might typically assume values such as $1-5\%$ of relevant failure probabilities. Anyhow, good choices of course depend on the general accuracy also induced by the sample number per subset N. If variation is high, we need not to be overly accurate in the approximation also.

In terms of implementation, we are responding to the challenges similar as described in Blandfort et al. (2019a), but choose an alternative way to cope with (C1-1) to achieve even higher efficiency regarding the total number of limit state evaluations (compare Section 2.1) and to guarantee termination of the algorithm. In Blandfort et al. (2019a), the number of total samples was considered instead. In this dissertation, we assume the following implementation, where we logically split the description with respect to the challenges as previously introduced (A1 replies to C1 etc.).

(A1) A1-1 Initialization by ordinary Subset Simulation: The dynamic variable is set to a low deterministic value $\mathbf{X}_k = y_1$, typically given by a quantile of $\mathbf{X}_k$ such that the domain is covered sufficiently well in every case (cf. Lemma 5.3.9). To cover a specific range of $\mathbf{X}_k$ values, we might also assume a uniform distribution on that range. The first limit state function is then defined by this deterministic value of $\mathbf{X}_k$. By the monotonicity assumption this corresponds to a rather likely failure with respect to the choice of $\mathbf{X}_k$. We then start with ordinary Subset Simulation until a subset $\mathbb{F}_i$ yields $b_i < b^*$. When reaching $b_i < b^*$, we set $b_i = b^*$ and continue with SuSI steps, changing $\mathbf{X}_k$ for the calculation of

successive failure probabilities. If the estimated failure probability under $\mathbf{X}_k = y_1$ is higher than a desired p_0, then a prior nested interval approach such as in Blandfort et al. (2019a) or alteration of y_1 helps to define an adequate initialization.

A1-2 Prediction Step, Scanning D_k for adequate successive y_{i+1} choices: This is the most challenging part of the algorithm since the conditional failure probability function q is unknown and can take an arbitrary form. Also a bias might be introduced if a specific choice y_{i+1} of $\mathbf{X}_k$ does not yield any failure. Naively changing $\mathbf{X}_k$ to another value for subsequent analysis then introduces a bias. Instead, we have to pre-sample values to guarantee a proper selection of y_{i+1} in advance so that at least one failure will occur with high probability. The proposed procedure is as follows. Given grid points $y_1, ..., y_i$, $i \in \mathbb{N}$, and their corresponding failure probability estimates $q(y_1), ..., q(y_i)$, we search for a $y_{i+1} \in D_k$ with $P(\mathbb{F}_{y_{i+1}}|\mathbb{F}_{y_i}) \approx p_0$. In all except for the first SuSI evaluation, this can be done by setting up an extrapolation function with grid point values as outputs and corresponding estimates as arguments. If we refer to this function by $f_{ext} : [0, 1] \to D_k$, then $f_{ext}\left(P(\mathbb{F}_{y_i})p_0\right)$ yields an estimate for a good choice of y_{i+1}. However, because this estimation could be bad (e.g. when q is highly non-linear), we propose to use a few samples of the previous subset for the possibility of testing for good y_{i+1} values. Assume such an evaluation with a candidate y_{i+1} yields $N_f \in \{0, ..., N_t\}$ fails out of $N_t \in \mathbb{N}$ samples. The fraction N_f/N_t is a guess of the intermediate probability $P(\mathbb{F}_{y_{i+1}}|\mathbb{F}_{y_i})$ at y_{i+1}. Indeed, this guess is demanded to imply that the probability of the true intermediate probability being included in a specific range of values $[p_l, p_u]$ is high. Fortunately, the true intermediate probability is known to be Beta distributed with parameters N_f and $N_t - N_f + 1$ by Proposition 4.2.11, so that we have

$$\begin{aligned} &P\left(P(\mathbb{F}_{y_{i+1}}|\mathbb{F}_{y_i}) \in (p_l, p_u)\right) \\ &= \text{Beta}(p_u N_t; N_f, N_t - N_f + 1) \\ &- \text{Beta}(p_l N_t; N_f, N_t - N_f + 1) \ . \end{aligned} \tag{5.2}$$

Defining a demanded probability for (5.2), such as $p_c = 90\%$,

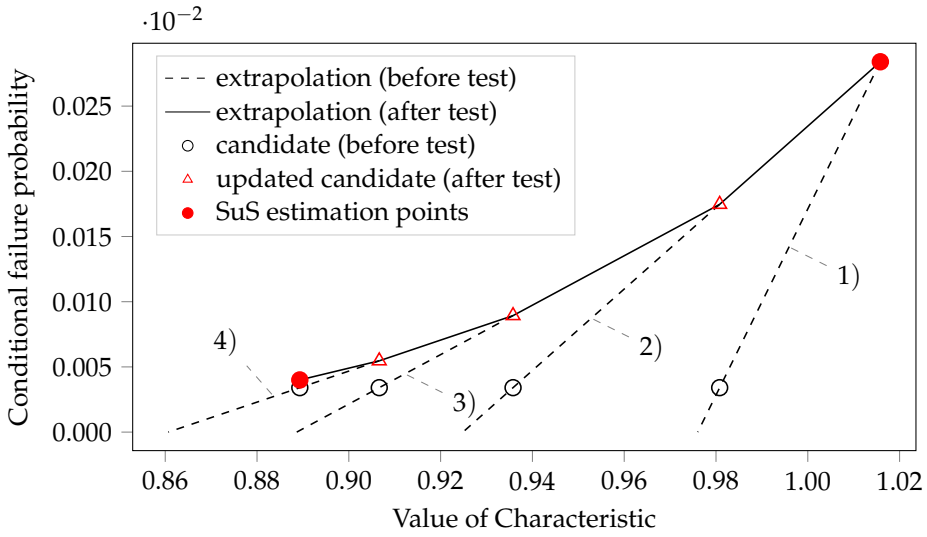


Figure 5.6: Blandfort et al. (2019a). Extrapolation step in the SuSI algorithm: Successive testing of candidate values until an acceptable $\mathbf{x}_k$ value, such that the conditional failure probability is likely to be contained in a selected interval of probabilities (p_l, p_u), is found. The final candidate is then taken for the next subset evaluation, defining y_{i+1}. In this illustration, contrasting Assumption 2.2.2, we have a decreasing $g(\mathbf{x}_1, ..., \mathbf{x}_d)$ with respect to $\mathbf{x}_k$.

yields a formal acceptance criterion for candidates. This test corresponds to a non-informative prior. A graphical illustration is given in Figure 5.6. We refer to this approach as 'version a' of the prediction step (compare Algorithm 2).

Although we still want to control our algorithm this way, we fortunately have shown in Section 4.1 that the coefficient of variation is insensitive with respect to p_0, if appropriate MCMC chain lengths are used, so that the interval (p_l, p_u) can be kept rather large to avoid many evaluations in the pre-steps. Then, we should adaptively define N dependent on N_f/N_t, such as $N = \frac{\log(p_0)}{\log(N_f/N_t)}$ to keep good efficiency. If several independent runs of SuSI are performed, previous results may be utilized for estimation of appropriate y_{i+1} instead (Algorithm 2, 'version b' of the prediction step). In this case, either the information can be used as a prior in

testing or the prediction step is completely replaced by the available information, not requiring any additional sampling efforts.

(A2) A2-1 Appropriate choices of y_1 and y_{m^*}: We propose to define y_1 as a quantile of the distribution function of $\mathbf{X}_k$. A straightforward choice is to set y_1 to the lowest $0.5\epsilon_{\max}$ quantile of all relevant distribution functions. This guarantees an error by evaluation smaller than $0.5\epsilon_{\max}$. If no distributions are given in advance, one can also set y_1 equal to any extraordinary small value in the domain of $\mathbf{X}_k$. Evaluation can be stopped when $\hat{q}(y_{i+1}) < 0.5\epsilon_{\max}$, having $y_{m^*} = y_{i+1}$ also resulting in an error less than $0.5\epsilon_{\max}$. If information about relevant distributions is given, we can also look at $(1 - \min_{F_k^j}\{F_k^j(y_1)\})\hat{q}(y_{i+1}) < 0.5\epsilon_{\max}$ instead, to avoid unnecessary evaluations.

A2-2 Simultaneous use of staircase and spline interpolation: We propose to use several interpolation methods so that one has not to rely on a single, possibly biased, result. In particular, staircase interpolation yields a guaranteed positive bias for upper and a negative bias for lower staircase case. This particularly allows to receive a result with guaranteed systematic overestimation of the failure probability by upper staircase interpolation, sticking to the principle of safety. On the other hand, for smooth conditional failure functions q, cubic spline interpolation yields approximately accurate results. If q is smooth, then this performs good in many situations. The bias can also be reduced by increasing p_0 or evaluation of additional grid points at important regions. Indeed, simulations suggest that it might even be sufficient to consider spline interpolation (compare Section 6.1) to reduce the bias.

In implementation, we notice the increased complexity of the algorithm. However, under our model assumptions, the algorithm is still general applicable and adaptively explores failure regions well in many cases. The full implementation is stated in Algorithm 2. Before demonstrating the full potential of SuSI, we want to be fair and not hold back drawbacks. These will be discussed in the following. Afterwards, we show why the additional implementation complexity is worth the effort.

Despite the increased gain of information by changing the limit state equation in each subset, there is also a disadvantage that must be considered

for a fair assessment of the method. To do so, we first need to define the resulting number of subset levels in SuSI that belong to SuS evaluations and those created according to the new procedure by altering $\mathbf{X}_k$ in successive subsets.

Definition 5.2.2. *The total number of subsets after termination of SuSI, is defined by m^*, where we refer to the subset level before the first SuSI-like step, i.e. the corresponding number of Monte Carlo and SuS subsets, as m_s. At grid point y_{m^*}, we achieve the minimal, and at y_1 the maximal, evaluated conditional failure probability.*

Remark 5.2.3 (Number of Evaluations)**.** *In SuSI the limit state function changes and it is unknown if the ordering of samples according to their limit state function values remains the same. Thus, in a new subset, all samples have to be evaluated again, even if they have been calculated in the previous subset. If not re-using seeds, this results in the following number of total evaluations for both the methods:*

$$\text{E-SuSI} \quad E_T = m_s N^{SuSI} + 2(m^* - m_s)N^{SuSI} = N^{SuSI}(2m^* - m_s)$$

$$\text{E-SuS} \quad E_T = mN^{SuS}$$

Under re-using seeds, the number of total evaluations in SuSI is increased even more, in particular for high p_0:

$$\begin{aligned}\text{Ep-SuSI} \quad E_T &\approx m_s(1-p_0)N^{SuSI} + (m^* - m_s)N^{SuSI}(2-p_0) \\ &= N^{SuSI}(m^*(2-p_0) - m_s)\end{aligned}$$

$$\text{Ep-SuS} \quad E_T \approx m_s(1-p_0)N^{SuS}$$

Comparing the two methods for same E_T thus often results in a smaller number of samples per subset in SuSI, $N^{SuSI} < N^{SuS}$. If we instead are rather interested in the number of generated samples which can be kept when $\mathbf{X}_k$ is changed in the evaluation or still know the ordering after changing $\mathbf{X}_k$, we have

$$\text{N-SuSI} \quad E_T = N_T = m^* N^{SuSI}$$

when not re-using seeds and

$$\text{Np-SuSI} \quad E_T = N_T \approx m^*(1-p_0)N^{SuSI}$$

otherwise, having an equal effort for SuSI and SuS-like evaluations. The additional evaluations of the prediction step (compare A1-2) are omitted due to several reasons. First, it is recommended to perform multiple independent simulation runs to derive an estimation result. In this case, however, we can also predict appropriate grid points by using the provided information of the first simulation run, not needing a prediction step anymore. Second, the sample number for the prediction step should be chosen small, so that it will become negligible if sample numbers are not very small. Third, the number of evaluations necessary for the predictions highly depends on the problem at hand and the implementation of the prediction step and it is therefore impossible to provide general results.

The higher number of evaluations in SuSI compared to SuS weakens its performance. At first, one might conclude that the computational demands of SuSI must thus be higher than in SuS. However, the number of SuSI evaluations is particularly higher for $\mathbf{X}_k$ that have a high impact on the limit state function. As a consequence, we have many additional evaluations for such $\mathbf{X}_k$ that also need many low level subsets for evaluation in the integration, which on the other hand increases efficiency because these generally provide a lower variation in their estimates. We will later see how those two effects might even cancel each other, resulting in a stable and good performance in many cases.

If the number of samples is the relevant measure for performance, SuSI even takes an approximately equal computational demand as SuS under same number of samples per subset N. Furthermore, if the ordering of the samples with respect to the limit state function is not changed in successive subsets, we have the same consequence. Also meta models might be a good idea to predict orderings and save computational time.

Lastly, we want to highlight that similar to SuS, it is advantageous to use several independent runs of SuSI as a result for reliability evaluation. This allows to use pre-calculated results for scanning of successive choices y_i, $i = 1, ..., m^*$, instead of doing an additional scanning step as in (A1-2). This not only reduces the number of evaluations, but might also yield better control of intermediate probabilities. Additionally, as the initial y_1 can be set to more reasonable values by the gained information, the number of total evaluations is further decreased. In addition, when several results are available, we may also use regression to derive a result. If the individual points then cover the domain finely, the bias can be decreased substantially.

Although often not explicitly mentioned, we propose to implement SuSI in the standard normal space (compare Section 2.1.1). This allows to refer to standard normal quantiles for coverage of the domain.

Remark 5.2.4. *In comparison to SuS, the main drawback of the SuSI algorithm might be its additional complexity in implementation. In particular the prediction step can be challenging, if one wants to cover all possible situations, including very extreme shapes of the conditional failure probability function q.*

Algorithm 2: Subset Simulation Interpolation (SuSI)

Result: Failure Probabilities $\hat{p}_f$ for all relevant distributions of $\mathbf{X}_k$

Initialization. Define
N (number of samples in each intermediate step),
N^{pred} (number of samples for prediction step),
p_0 (intermediate subset probability for SuS),
(p_l, p_u) (interval for admissible probabilities of intermediate subsets for SuSI),
$\mathbf{X}_k$ (stochastic property considered for SuSI),
F_k^* ('worst' relevant distributions of $\mathbf{X}_k$, see step A2-1. and step A2-2.),
$\epsilon_{\max}$ (maximal admissible domain error)

A2-1. SuS Subsets.
$y_1 := \min_{F_k^*}\{c_{0.5\epsilon_{\max},F_k}\}$ the lowest $\frac{1}{2}\epsilon_{\max}$ quantile
set limit state $g(\mathbf{X}|\mathbf{X}_k = y_1)$,
evaluation with Monte Carlo,
$i = 1$
while *subset threshold* $b_i > b^*$ **do**
  continue ordinary Subset Simulation, subset level $i = i + 1$
end

A2-2. SuSI Subsets.
save starting subset $m_s = i$
while $(1 - F_k^*(y_i))\hat{q}(y_i) > \frac{1}{2}\epsilon_{\max}$ **do**
  create N samples of the previous subset $\{\mathbf{x} : g(\mathbf{x}|\mathbf{X}_k = y_i) < b^*\}$ by MCMC (as in SuS)
  Prediction Step:
    a) find candidate y_{i+1} s.t. $\hat{p}_{i+1} \in (p_l, p_u)$ is likely by few samples $N^{pred} \ll N$
    b) if a previous SuSI result is availble, use A2-4 to find y_{i+1} so that $\hat{q}(y_{i+1})/\hat{q}(y_i) = \frac{1}{2}(p_l + p_u)$
  evaluate $g(\mathbf{x}|\mathbf{X}_k = y_{i+1}) < b^*$ for all samples and compute $\hat{p}_i = \hat{P}(\mathbb{F}_{y_{i+1}}|\mathbb{F}_{y_i})$
  $i = i + 1$
end

A2-3. Simulation Result.
grid points $y_{m_s}, ..., y_{m^*}$
estimated conditional failure probabilities: $\hat{q}(y_i) = \prod_{j=1}^{m_s+i} \hat{p}_j$, $i = m_s, ..., m^*$

A2-4. Interpolation.
choose interpolation method and create a continuous result $\hat{q}(y)$ for $y \in D_k$, using the grid point results $\hat{q}(y_{m_s}), ..., \hat{q}(y_{m^*})$

A2-5. Evaluating failure probabilities for any f_k, by numerical interpolation.
$\hat{p}_f = \int_{D_k} \hat{q}(y) f_k(y) dy$

5.3 Statistical Analysis

We have introduced the idea and implementation of the new algorithm. Now, we explore the theoretical basis for it. We are going to answer the question:

What is the computational demand of applying SuSI in comparison to ordinary SuS in a static reliability analysis?

Note, we compare a single run of SuS and SuSI here, whereas SuSI was developed for evaluation of several model constellations n, since it only requires one simulation run, independent of n. Therefore, in case of similar efficiency for a single evaluation, we can conclude a great performance of SuSI, providing computational demands as low as a static evaluation for dynamic reliability analysis. Even if not comparing efficiency of SuSI and SuS directly, it enables an intuitive access to the computational demands of applying SuSI, relating it to the demands of a popular algorithm in a static reliability analysis. The results on statistical analysis of splines could be considered hard to interpret. However, the proofs already implicitly offer a better understanding, show some proof strategies for the special setting we are confronted with and furthermore aim at providing a contribution to statistical analysis of splines with Gaussian distributed knot values.

5.3.1 Preliminaries

First, it is plausible to assume the statistical properties of single Subset Simulation estimators to also hold for the individual grid point estimators produced by the altered SuS-like steps in SuSI because the applied procedure for their creation is basically the same. Anyhow, in contrast, the intermediate subset probability estimators in SuSI are unbiased while SuS steps are asymptotically unbiased with rate $O(1/N)$ as discussed in Section 4.2. Unbiasedness in SuSI is due to pre-definition of the threshold by the prediction step in SuSI (cf. fixed levels method in Cérou et al. (2012)). However, we also start with ordinary Subset Simulation often when applying SuSI so that estimation results also rely on SuS steps in many cases.

To simplify analysis, we assume Gaussian distributed intermediate probability estimators $\hat{P}_i$ with mean value $p_{r,i} = p_0$ (Assumption 5.3.1) for all subsets $\mathbb{F}_i$ (and $\mathbb{F}_{y_i}$), $i = 1, ..., m^*$. This is the traditional approach for analyzing Subset Simulation (Au and Beck (2001b)) and clearly results in many advantages for further analysis. This assumption relates to a priori chosen

thresholds with true probability p_0 and large sample number N. Indeed, this will not suit reality exactly (compare Section 4.2), but is sufficient in our case since we focus on general properties and not on details or small sample numbers. Furthermore, the simplifications allow to derive results that are much easier to interpret, even under the increased complexity by interpolation. Linear combinations of Gaussian distributed random variables, which are then the basis of our analysis, are rather easy to handle.

Assumption 5.3.1. *The intermediate subset probability estimators, in SuS and SuSI, for simplicity writing $\hat{P}_i$, $i = 1, 2, ...$ jointly for both approaches, are assumed Gaussian distributed and unbiased*

$$\hat{P}_i = p_{r,i} + \sigma_i Z_i$$

for standard normally distributed random variables $Z_1, Z_2, ...$ with variance

$$\sigma_i^2 = \frac{p_{r,i}(1 - p_{r,i})}{N}(1 + \gamma_i)$$

in $O\left(\frac{1}{N}\right)$ where γ_i accounts for conversion of dependent samples to an effective number of independent samples. According to Remark 4.2.17, we further assume $Z_1, Z_2, ...$ to be independent.

Note, the simplified approach as by Assumption 5.3.1 suits SuSI better with respect to the a priori selection of thresholds. Our aim is to analyze SuSI with SuS as a benchmark, looking at relative behavior. The assumptions and approximations are made for both the methods, so that our simplifications should not change the results significantly if both methods are affected similarly. Next, we define a substantial simplification of the statistical analysis that is kept throughout the whole section and also corresponds to a typical assumption in analysis of SuS (compare Assumption 3.1.6).

Remark 5.3.2. *Terms of order $O(1/N)$ are neglected.*

The assumption of neglecting $O(1/N)$ terms, under dependent intermediate subset probability estimators, results in the same consequences as independent intermediate subset probability estimators in many parts of the analysis. So we could, instead of proceeding as stated in Remark 5.3.2, also often use independence given by Assumption 5.3.1 to proceed with the analysis. However, we prefer the asymptotic considerations most of the time

because Assumption 5.3.1 is more plausible under high N and details as derived in Section 4.2 are neglected. Also this suffices our demands here as we are interested in a general comparison of SuS and SuSI where conclusions should not be substantially altered unless the sample number is very small. Nevertheless, it is important to keep in mind that all results in this section are asymptotic results up to order $O(1/N)$ so that they hold asymptotically, thus only for large N.

Additionally, we also shorten the notation and rewrite the monotonicity assumption by the corresponding conditional failure probability function as given in the following Remark.

Remark 5.3.3 (Definition and Monotonicity of the Conditional Failure Probability q)**.** *By monotonicity Assumption 2.2.2, the conditional failure probability* $q : D_k \to [0,1]$,

$$q(y) := P(g(\mathbf{X}) < b^* | \mathbf{X}_k = y) = \int_D f_{\mathbf{X}}(\mathbf{x}) \mathbf{1}_{\{g(\mathbf{x}) < b^* | \mathbf{x}_k = y\}} d\mathbf{x}$$

is monotone decreasing with respect to $\mathbf{x}_k$*. We will sometimes write short* $q_w = q(y_w)$*,* $w \in \{1, ..., m^*\}$ *with respect to the grid points* $y_1, ..., y_{m^*}$ *in SuSI for ease of notation and write an uppercase letter* $\hat{Q}$ *for the corresponding stochastic variable.*

Because of the assumption of Gaussian distributed estimators and predefined thresholds, the following conclusions closely follow the ones given in Au and Beck (2001b) for ordinary SuS.

Proposition 5.3.4. *Under Gaussian distributed intermediate probability estimators (Assumption 5.3.1), the estimation error* $\mathcal{E}_w := \hat{Q}(y_w) - q(y_w)$ *of the failure probability at grid point* y_w*,* $w = 1, ..., m^*$ *in SuSI is given by*

$$\mathcal{E}_w = \sum_{i=1}^{w} \left(\sigma_i \prod_{j \neq i} p_{r,j} \right) Z_i$$

for $Z_1, ..., Z_w$ *standard normally distributed* $\mathcal{N}(0,1)$*,* σ_i^2 *the variance of the intermediate subset probability estimators* $\hat{P}_i$*. Consequently,* $\hat{Q}(y_w)$ *is asymptotically unbiased.*

Proof. By $p_{r,i} = P(\mathbb{F}_{i+1} | \mathbb{F}_i)$, the true intermediate subset probability of subset $i = 1, ..., m^*$, we have $q(y_w) = \prod_{i=1}^{w} p_{r,i}$ and equivalently for the stochastic

variable $\mathcal{E}_w$,

$$\mathcal{E}_w = \hat{Q}(y_w) - q(y_w) = \prod_{i=1}^{w} \hat{P}_i - \prod_{i=1}^{w} p_{r,i} \ .$$

Let $Z_1, ..., Z_w$ be standard normally distributed random variables and $\sigma_i^2 = O(1/N)$ (see Assumption 5.3.1) the variance of $\hat{P}_i$. By Assumption 5.3.1, we get

$$\begin{aligned}\mathcal{E}_w &= \prod_{i=1}^{w}(p_{r,i} + \sigma_i Z_i) - \prod_{i=1}^{w} p_{r,i} \\ &= \sum_{i=1}^{w} \sigma_i Z_i \prod_{j\neq i} p_{r,j} + \sum_{i=1}^{w}\sum_{j>i}^{w} Z_i Z_j \sigma_i \sigma_j \prod_{l\neq i,j} p_{r,l} + ...\end{aligned}$$

From $\sigma_i = O(1/\sqrt{N})$ it follows

$$\mathcal{E}_w = \sum_{i=1}^{w} \sigma_i Z_i \prod_{j\neq i} p_{r,j} + O(1/N) \ .$$

Neglecting terms of order $O(1/N)$ according to Remark 5.3.2 yields the main result. Unbiasedness then follows directly by $E[Z_1] = ... = E[Z_w] = 0$. □

Next, we can explore statistical properties of the derived error in the SuS-like estimators, but first add an assumption for ease of notation and analysis. At the same time, we formalize the assumption of identical stochastic properties of SuS-like estimators in SuS and SuSI.

Assumption 5.3.5 (Homogeneous Subsets Assumption)**.** *We assume the intermediate subset probability estimators $\hat{P}_i = P(\mathbb{F}_{i+1}|\mathbb{F}_i)$ by SuS and SuSI to be homogeneous, i.e. $p_{r,i} = p_0$ and $\sigma_i^2 = \sigma^2$ for all $i = 1, 2, ..., m^*$. In addition, we also assume that the local variances σ^2 of the estimators in SuS and SuSI coincide under identical intermediate probability p_0 and number of samples N per subset. This corresponds to assuming a constant γ_i in Assumption 5.3.1 for all subsets and independent of SuS or SuSI type steps. Since there is no systematic difference between the two methods, we assume $\gamma_i = 0$ (cf. Assumption 3.1.10) in the following for ease of notation, corresponding to perfect sampling. Later we will also comment on the effects of non-perfect sampling $\gamma_i > 0$.*

Although the SuSI estimators are not derived in the exact same manner as in SuS, as the limit state might be exposed to more general changes in every subset, the Homogeneous Subset Assumption is still plausible as the creation procedure is very similar. Nevertheless, the approximately constant $p_{r,i} = p_0$ in each subset should better be fulfilled in SuS, at least when not performing several independent SuSI runs so that previous information can be utilized for selecting appropriate grid points in SuSI.

Proposition 5.3.6. *Under Assumption 5.3.5, the error* $\mathcal{E}_w = \hat{Q}(y_w) - q(y_w)$ *of the grid point estimator has variance* $\mathrm{Var}(\mathcal{E}_w)$ *and covariances* $\mathrm{Cov}(\mathcal{E}_w, \mathcal{E}_l)$*, for* $w, l \in \{1, ..., m^*\}$*, as given by*

(i) $\mathrm{Var}(\mathcal{E}_w) = w\sigma^2 p_0^{2(w-1)}$,

(ii) $\mathrm{Cov}(\mathcal{E}_w, \mathcal{E}_l) = \mathrm{Var}(\mathcal{E}_w) p_0^{l-w}$ *for grid points* y_w, y_l *with* $w \leq l$.

Proof. We start with the representation in Proposition 5.3.4. Using unbiasedness of the estimator, the variance of the grid point estimators is given by

$$\mathrm{Var}(\mathcal{E}_w) = \sum_{i=1}^{w} \left(\sigma_i^2 \prod_{j \neq i} p_{r,j}^2 \right) .$$

Now Assumption 5.3.5 yields

$$\mathrm{Var}(\mathcal{E}_w) = w\sigma^2 p_0^{2(w-1)} ,$$

which proves the first part.

For the second part, again by Proposition 5.3.4, we have

$$\mathcal{E}_w = \sum_{i=1}^{w} \left(\sigma_i \prod_{j \neq i} p_{r,j} \right) Z_i$$

for some standard normally distributed Z_i, $i = 1, ..., w$. The covariance is thus given by

$$\begin{aligned} \mathrm{Cov}(\mathcal{E}_w, \mathcal{E}_l) &= E\left[\left(\sum_{i=1}^{w} \sigma_i Z_i \prod_{q \neq i} p_{r,q} - 0 \right) \left(\sum_{j=1}^{l} \sigma_j Z_j \prod_{k \neq j} p_{r,k} - 0 \right) \right] \\ &= E\left[\left(\sum_{i}^{w} \sum_{j}^{l} \sigma_i \sigma_j E\left[Z_i Z_j \right] \prod_{q \neq i}^{w} p_{r,q} \prod_{k \neq j}^{l} p_{r,k} \right) \right] . \end{aligned}$$

By independence of $Z_i, Z_j, i \neq j$ (compare Assumption 5.3.1), this is equal to

$$\sum_{i=1}^{w} \sigma_i^2 \prod_{q \neq i}^{w} p_{r,q} \prod_{k \neq i}^{l} p_{r,k} \ .$$

Now using the homogeneous subsets assumption (Assumption 5.3.5) yields

$$\mathrm{Cov}(\mathcal{E}_w, \mathcal{E}_l) = \sum_{i=1}^{w} \sigma^2 p_0^{(w-1)+(l-1)} = w\sigma^2 p_0^{(w-1)+(l-1)} \ .$$

Comparison with the variance of $\mathcal{E}_w$

$$\mathrm{Var}(\mathcal{E}_w) = w\sigma^2 p_0^{2(w-1)}$$

allows the reformulation

$$\mathrm{Cov}(\mathcal{E}_w, \mathcal{E}_l) = \mathrm{Var}(\mathcal{E}_w) p_0^{l-w}$$

where $l - w \geq 0$. □

Next, we make two technical assumptions and simplify notation and analysis without much loss of generality.

Assumption 5.3.7. *The conditional failure probability q is assumed to be continuous.*

Also, we again use Assumption 3.1.5, assuming that p_f can be written as $p_f = p_0^m$, m the integer number of subsets in SuS with conditional probabilities equal to $p_0 \in (0, 1)$. If Assumption 3.1.5 is violated, m would define the final subset level in SuS with $\hat{p}_m > p_0$. Then however notation is less handy. In the considered case, SuS is also most efficient as the number of samples per subset N is chosen perfectly in the last subset, which it is not in general. The number of necessary subset levels for sufficient accuracy in SuSI may generally differ to that of SuS. This difference is due to the fact that SuSI integrates over the conditional failure probabilities where values smaller than p_0^m might be necessary to get the demanded accuracy, requiring $m^* > m$. On the other hand it might also be the case that only values greater than p_0^m are relevant for evaluation, then needing less subsets in SuSI, $m^* < m$ under same p_0.

Remark 5.3.8. *Although SuSI utilizes ordinary SuS up to level m_s before beginning with SuSI steps in general, we may mostly ignore this in our analysis as it does not affect the derived results in many cases. Still note that usually the first subsets all belong to the same choice of $\mathbf{X}_k$ so that an interpolation between those is technically impossible. An exception where we can not ignore this is the number of evaluations E_T, where we have a significant difference between using SuS or SuSI steps (compare Remark 5.2.3).*

In addition to the simplification in Remark 5.3.8, we neglect the error by an imperfect domain coverage for ease of notation and assume it is small enough by a proper selection of y_1 and y_{m^*}, where Lemma 5.3.9 shows that this error can anyways be reduced arbitrarily by a suitable construction of the subsets.

Lemma 5.3.9. *For every $\epsilon > 0$, we find $y_1, y_{m^*} \in D_k$ so that the domain $[y_1, y_{m^*}]$ for approximation in SuSI allows to calculate the probability of failure accurately up to error ϵ:*

$$\int_{-\infty}^{y_1} f_k(y)q(y)dy + \int_{y_{m^*}}^{\infty} f_k(y)q(y)dy \leq F_k(y_1) + (1 - F_k(y_{m^*}))q(y_{m^*}) < \epsilon \ .$$

Proof. This is straightforward by continuity and by boundedness of non-degenerate pdfs and the natural boundary $[0, 1]$ of the conditional failure probability q. □

In the next sections, we analyze two interpolation methods. In both cases, we can understand the results by interpolation as a linear combination of SuS-like produced estimates. A first question one might pose is why we use interpolation and not regression when trying to approximate the conditional failure probability function q.

Remark 5.3.10 (Preferring Interpolation to Regression)**.** *Preferring interpolation to regression has the purpose of keeping as much confidence as possible on the failure probability estimation by the algorithm. There are already many sources of uncertainty such as unknown effective number of samples, a potential bias or the question of ergodicity in Subset Simulation. Thus, we do not want to create an algorithm that is even worse interpretable than Subset Simulation is already. Indeed we even demonstrate how SuSI can help to provide evidence on SuS results. Having the opportunity to rely directly on the SuS-like generated samples by interpolation promises a possibly tractable*

increase in uncertainty as we have similarity of the grid point estimators to classical Subset Simulation estimators, allowing to build up a theoretical foundation directly on SuS. Furthermore, the values of the given discrete set of estimates by SuSI are differing a lot so that it might not be a good idea to smooth the regression function at the cost of not hitting the estimates directly. Indeed, these point estimates are the best approximately unbiased information given on the unknown shape of q.

In contrast to Remark 5.3.10, if we have many repetitive calculations of SuSI and thereby many estimates for one or several close grid point values, then it can be beneficial to use regression. In this case, efficiency might be improved without adding relevant uncertainty on the bias at the grid points. In particular it should be possible to erase the bias by a dense coverage of the domain of $\mathbf{X}_k$ if many points are available. Nevertheless, before proceeding with such an analysis, it is necessary to better understand SuSI in general and also to be capable of performing SuSI by only one or a few independent simulation runs of it. This also allows better comparison of statistical properties with respect to SuS.

In both cases, interpolation and regression, the result is generated by a weighted sum of SuS-like estimates. We thus need to derive properties of such linear combinations of correlated SuS-like estimators where we study interpolation approaches. To proceed by SuSI, one needs to choose a concrete interpolation method. We analyze a staircase approximation as well as cubic spline interpolation. Estimation of failure probabilities by such interpolation methods certainly requires to consider the thereby introduced bias. In comparison to ordinary SuS, we then have to consider both, bias and coefficient of variation jointly to reasonably judge the efficiency of the new algorithm.

5.3.2 Staircase Approach

We first analyze SuSI with staircase interpolation.

Definition 5.3.11 (Staircase Interpolation)**.** *The upper and lower staircase approximations $s_u, s_l : D_k \to [0,1]$ of the conditional failure probability function q are given by*

$$s^{\mathrm{u}}(y) := q(y_i) \text{ and } s^{\mathrm{l}}(y) := q(y_{i+1}) \text{ for } y \in [y_i, y_{i+1}) \ ,$$

where $i \in \{1, ..., m^* - 1\}$. To cover the whole domain, we also define

$$s^{\mathrm{u}}(y) := 1,\ s^{\mathrm{l}}(y) := q(y_1) \text{ for } y < y_1$$

and

$$s^{\mathrm{u}}(y) := q(y_{m^*}),\ s^{\mathrm{l}}(y) := 0 \text{ for } y \geq y_{m^*}$$

which is an upper boundary by the natural probability boundaries $q(y) \in [0,1]$ *for all* $y \in D_k$*. By Lemma 5.3.9 the error by not fully covering the domain can be made arbitrarily small by appropriate choice of* y_1 *and* y_{m^*}*.*

The staircase interpolation for approximation of the conditional failure function is important due to the following reason.

Remark 5.3.12 (Choosing the Staircase Approach)**.** *As the conditional failure probability function* q *is in general unknown, only known to be monotone, the only function that guarantees a non-negative bias, avoiding systematic underestimation of the failure probability is the upper staircase function. For every other approximation method, it is not clear if the estimation has a negative bias and therefore does converge to a value lower than the true failure probability for high* N*, systematically underestimating the real failure probability. An additional point is that the staircase function allows for a better comprehensible theoretical analysis. As SuSI is a newly proposed approach, we should start with a strong basis where analysis remains simple enough to have a good understanding of the mechanisms. It is also straightforward, although perhaps lengthy in analysis, to adapt the staircase results to more sophisticated interpolation methods as shown in the next section.*

Our objective is to measure efficiency by comparison of the MSE for same number of function evaluations E_T (compare Definition 2.1.2). To do so, we use the well known relation

$$\mathrm{MSE}(\hat{P}_f) = \mathrm{Var}(\hat{P}_f) + \mathrm{B}(\hat{P}_f)^2$$

and separately analyze bias ($\mathrm{B}(\hat{P}_f)$) and variance ($\mathrm{Var}(\hat{P}_f)$) in the following, afterwards putting together both the results. To clarify the method used for estimation, we write $\hat{P}_f^{\mathrm{sus}}$ for classical Subset Simulation and $\hat{P}_f^{\mathrm{u}}$ and $\hat{P}_f^{\mathrm{l}}$ for Subset Simulation Interpolation with upper and lower staircase interpolation, respectively. Note, these estimators always correspond to specific choices of parameters, samples per subset N and intermediate subset probability p_0.

Theorem 5.3.13 (Bias of SuSI)**.** *The relative bias of the estimator in upper staircase interpolation is given by*

$$\frac{\mathrm{B}(\hat{P}_f^{\mathrm{u}})}{p_f} = c_u(p_0)\frac{1-p_0}{p_0} \tag{5.3}$$

for some $c_u(p_0) \in [0,1]$. *Correspondingly the relative bias of the estimator under lower staircase interpolation is given by*

$$\frac{\mathrm{B}(\hat{P}_f^{\mathrm{l}})}{p_f} = -c_l(p_0)\frac{1-p_0}{p_0} \tag{5.4}$$

for some $c_l(p_0) \in [0,1]$.

Proof. We proceed in two steps, first deriving the boundaries of the bias and then using the intermediate value theorem to prove the claim. We only show the upper staircase case as the lower staircase case is proven alike.

In SuSI$^{\mathrm{u}}$ the estimated conditional probability of failure of $y \in [y_i, y_{i+1}] \subseteq D_k, i \in \{1, ..., m^* - 1\}$, is given by the estimate of the upper staircase function $\hat{S}^{\mathrm{u}}(y) = \hat{Q}(y_i)$. By unbiasedness of the grid point estimators (Proposition 5.3.4), we thus have $E\left[\hat{S}^{\mathrm{u}}(y)\right] = E\left[\hat{Q}(y_i)\right] = q(y_i)$. Similarly, we have $E\left[\hat{S}^{\mathrm{l}}(y)\right] = E\left[\hat{Q}(y_{i+1})\right] = q(y_{i+1})$ for the lower staircase function approximation. The bias of the upper staircase estimation with respect to the whole interval $[y_i, y_{i+1}]$ is given by

$$\mathrm{B}_i(\hat{P}_f^{\mathrm{u}}) = \int_{y_i}^{y_{i+1}} f_k(y)(q(y_i) - q(y))dy \ .$$

By monotonicity of the conditional failure function $q(y)$, we have $q(y_{i+1}) = s^{\mathrm{l}}(y) \leq q(y) \leq s^{\mathrm{u}}(y) = q(y_i)$ and by Assumption 5.3.5 $q(y_{i+1}) = p_0 q(y_i)$ so that

$$\mathrm{B}_i(\hat{P}_f^{\mathrm{u}}) \leq \int_{y_i}^{y_{i+1}} f_k(y)(q(y_i) - p_0 q(y_i))dy \ .$$

Reformulation yields

$$\mathrm{B}_i(\hat{P}_f^{\mathrm{u}}) \leq (1-p_0)\left(\int_{y_i}^{y_{i+1}} f_k(y)q(y)dy + \int_{y_i}^{y_{i+1}} f_k(y)(q(y_i) - q(y))dy\right) \ .$$

Thus the bias over the whole domain D_k is bounded by

$$\mathrm{B}(\hat{P}_f^{\mathrm{u}}) \leq (1-p_0)\sum_{i=1}^{m^*-1}\left(\int_{y_i}^{y_{i+1}} f_k(y)q(y)dy + \int_{y_i}^{y_{i+1}} f_k(y)(q(y_i) - q(y))dy\right)$$

up to approximation error as in Lemma 5.3.9, which is neglected. This is just the upper staircase approximation multiplied by $(1 - p_0)$ so that

$$\mathrm{B}(\hat{P}_f^{\mathrm{u}}) \leq (1 - p_0)(p_f + \mathrm{B}(\hat{P}_f^{\mathrm{u}})) \ .$$

Reordering yields the upper boundary

$$\frac{\mathrm{B}(\hat{P}_f^{\mathrm{u}})}{p_f} \leq \frac{1 - p_0}{p_0}$$

where the boundary is also known to be positive by definition of s^{u} and monotonicity of q.

Now the intermediate value theorem on function $f : [0, 1] \to \left[0, \frac{1-p_0}{p_0}\right]$ with $f(c) := c\frac{1-p_0}{p_0}$ yields the result.

□

Proposition 5.3.14. *The relation of the factors for upper and lower staircase bias in Theorem 5.3.13 is given by*

$$1 = \frac{c_l(p_0)}{p_0} + c_u(p_0) \ .$$

Proof. Write $\hat{P}_f^{\mathrm{u}}$ and $\hat{P}_f^{\mathrm{l}}$ for the upper and lower staircase estimators by SuSI. The definition of the bias gives

$$p_f = E[\hat{P}_f^{\mathrm{l}}] - \mathrm{B}(\hat{P}_f^{\mathrm{l}}) \ .$$

By construction we also know the relation between upper and lower staircase expected values of estimation to be given by

$$p_f = p_0 E[\hat{P}_f^{\mathrm{u}}] - \mathrm{B}(\hat{P}_f^{\mathrm{l}}) = p_0(p_f + \mathrm{B}(\hat{P}_f^{\mathrm{u}})) - \mathrm{B}(\hat{P}_f^{\mathrm{l}}) \ .$$

Theorem 5.3.13 then yields

$$p_f = p_0 \left(p_f + p_f c_u \frac{1 - p_0}{p_0} \right) + p_f c_l \frac{1 - p_0}{p_0} \ .$$

Crossing out p_f

$$1 = p_0 \left(1 + c_u \frac{1 - p_0}{p_0} \right) + c_l \frac{1 - p_0}{p_0}$$

and reformulating

$$1 - p_0 = \frac{1 - p_0}{p_0}(p_0 c_u + c_l)$$

yields

$$1 = c_u + \frac{c_l}{p_0} \ .$$

□

Proposition 5.3.14 also indicates that the more biased a lower staircase approximation is, the less biased the upper will be and vice versa. This is straightforward as q takes values between upper and lower staircase functions by monotonicity. The factor $\frac{1}{p_0}$ comes from the fact that a smaller lower staircase bias corresponds to a smaller p_f. This again affects the relative bias with respect to p_f.

Remark 5.3.15. *Upper and lower staircase bias yield the highest possible positive and negative bias for monotone interpolation, respectively. Thus the bias under every monotone interpolation method in SuSI can be written by either Equation 5.3 or Equation 5.4 by choosing appropriate c_u or c_l values. However, it might miss information on the behavior of the bias such as decrease of a specific order with respect to the length of the grid point intervals in spline interpolation. Even in staircase interpolation, the factors in the equations are not fixed and change by the placement of the grid points or equivalently p_0. Still, we will treat the factors as constants c_u and c_l, not further analyzing their dependency on p_0, in the following. This is due to the fact that they are generally not traceable, depending on the unknown real function q.*

Remark 5.3.16. *Using the average of upper and lower staircase function for interpolation, the bias follows*

$$\frac{|\mathrm{B}(\hat{P}_f^{\mathrm{average}})|}{p_f} \leq \frac{1}{2} \cdot \frac{1 - p_0}{p_0} \ ,$$

since one bias is positive and the other is negative. Moreover, this allows a reduction of the absolute value of the bias boundary, but may also result in a positive or a negative bias. In other words, we expect to perform better for many functions at the cost of uncertainty in the sign of the bias.

Next, after proving as preparation the following claim, we dedicate ourselves to the variation of estimations by SuSI.

Lemma 5.3.17. *Under Assumption 3.1.5 and for given pdf f_k and $\mathcal{F}_1, ..., \mathcal{F}_{m^*-1}$ defined as $\mathcal{F}_i := \int_{y_i}^{y_{i+1}} f_k(y)dy$, $i = 1, ..., m^* - 1$ and corresponding grid points $y_1, ..., y_{m^*}$ by SuSI, we have*

$$\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \frac{i}{m} p_0^{2i} + 2 \sum_{1 \le i < j \le m^*-1} \mathcal{F}_i \mathcal{F}_j \frac{i}{m} p_0^{i+j}$$

$$= C_u \cdot p_0^{2m} \left(1 + p_0^{-m} \mathrm{B}(\hat{P}_f^{\mathrm{u}})\right)^2$$

for some $C_u \in [\frac{1}{m}, \frac{m^*-1}{m}]$ and $\mathrm{B}(\hat{P}_f^{\mathrm{u}})$ the bias of the upper staircase approximation. Similar, we have for the lower staircase approximation

$$\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \frac{i+1}{m} p_0^{2i+2} + 2 \sum_{1 \le i < j \le m^*-1} \mathcal{F}_i \mathcal{F}_j \frac{i+1}{m} p_0^{i+j+2}$$

$$= C_l \cdot p_0^{2m} \left(1 + p_0^{-m} \mathrm{B}(\hat{P}_f^{\mathrm{l}})\right)^2$$

for some $C_l \in [\frac{2}{m}, \frac{m^*}{m}]$ and $\mathrm{B}(\hat{P}_f^{\mathrm{l}})$ the bias of the lower staircase approximation.

Proof. For the upper staircase approximation, we have

$$E\left[\hat{P}_f^{\mathrm{u}}\right] = \sum_{i=1}^{m^*-1} \int_{y_i}^{y_{i+1}} f_k(y) q(y_i) dy = \sum_{i=1}^{m^*-1} \mathcal{F}_i q(y_i) \tag{5.5}$$

and

$$E\left[\hat{P}_f^{\mathrm{u}}\right] = p_f + \mathrm{B}(\hat{P}_f^{\mathrm{u}}) \tag{5.6}$$

where $\mathrm{B}(\hat{P}_f^{\mathrm{u}})$ is the bias of the upper staircase estimation. Combining (5.5) and (5.6) yields

$$\sum_{i=1}^{m^*-1} \mathcal{F}_i q(y_i) = p_f + \mathrm{B}(\hat{P}_f^{\mathrm{u}}) \ .$$

Then squaring both sides and using $q(y_i) = p_0^i$ by upper staircase approximation gives

$$\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 p_0^{2i} + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j p_0^{i+j} = p_f^2 + 2p_f \mathrm{B}(\hat{P}_f^{\mathrm{u}}) + (\mathrm{B}(\hat{P}_f^{\mathrm{u}}))^2 \ .$$

Replacing p_f by p_0^m results in

$$\underbrace{\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 p_0^{2i} + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j p_0^{i+j}}_{=:L} = \underbrace{p_0^{2m} \left(1 + p_0^{-m} \mathrm{B}(\hat{P}_f^{\mathrm{u}})\right)^2}_{=:R} \ . \tag{5.7}$$

Now we want to look at

$$L^* := \sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \frac{i}{m} p_0^{2i} + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j \frac{i}{m} p_0^{i+j}$$

which is then, due to $\frac{i}{m} \in \{\frac{1}{m}, \frac{2}{m}, ..., \frac{m^*-1}{m}\}$, bounded by

$$\frac{1}{m} L \le L^* \le \frac{m^*-1}{m} L \ .$$

Thus, by (5.7) we may derive boundaries for L^* as

$$\frac{1}{m} R \le L^* \le \frac{m^*-1}{m} R \ .$$

Now the intermediate value theorem with respect to the interval $[\frac{1}{m}, \frac{m^*-1}{m}]$ yields the claim. The lower staircase case is shown accordingly. □

Theorem 5.3.18 (Variance of SuSI)**.** *Let M be the relation of sample numbers, i.e. $M = \frac{N^{SuS}}{N^{SuSI}}$. Under Assumption 5.3.5 and Assumption 3.1.5, the relation of the variance of estimated failure probabilities by upper staircase in SuSI and classical SuS variance with same p_0 and number of total evaluations E_T is given by*

$$\frac{\mathrm{Var}(\hat{P}_f^{\mathrm{u}})}{\mathrm{Var}(\hat{P}_f^{\mathrm{sus}})} = MC_u \left(1 + c_u \frac{1-p_0}{p_0}\right)^2$$

for some $c_u \in [0,1]$ as in Proposition 5.3.14 and some $C_u \in [\frac{1}{m}, \frac{m^*-1}{m}]$ as in Lemma 5.3.17. In the case of lower staircase approximation, we correspondingly have

$$\frac{\mathrm{Var}(\hat{P}_f^{\mathrm{l}})}{\mathrm{Var}(\hat{P}_f^{\mathrm{sus}})} = MC_l \left(1 - c_l \frac{1-p_0}{p_0}\right)^2$$

for $c_l \in [0,1]$ with $c_l = (1-c_u)p_0$ by Proposition 5.3.14 and some $C_l \in [\frac{2}{m}, \frac{m^*}{m}]$.

If seeds of successive subsets are not re-used for evaluation in the next subset, M is given by $\frac{2m^*-m_s}{m}$.

Proof. We proceed as follows: In a first step, we write $\hat{P}_f^{\mathrm{u}}$ as a linear combination of SuS-like produced estimates. Weights are derived by the interpolation method as well as the probability distribution of $\mathbf{X}_k$. The SuS estimates are given by the product of the estimated intermediate probabilities of the subsets up to a specific subset. Note, in SuSI those SuS estimates are identified as estimated conditional failure probabilities at the grid points $\hat{q}(y_i), i = 1, ..., m^* - 1$. Then we use the formula for variances and covariances in Proposition 5.3.6 with sample numbers per subset as derived in Remark 5.2.3. This allows to compare the variances of SuS and individual grid point SuSI estimators with respect to a fixed number of total limit state calls E_T. This result is then used to derive the variance of the weighted SuS estimators in $\hat{P}_f^{\mathrm{u}}$ in relation to the single SuS estimator in SuS.

Interpolation: Weighted SuS Estimates Write $\mathcal{F}_i := \int_{y_i}^{y_{i+1}} f_k(y)dy$ and, as above, $\mathcal{E}_i = \hat{Q}(y_i) - q(y_i)$ the estimation error at grid points $y_i, i = 1, ..., m^* - 1$ by SuSI. The variance of the upper staircase estimator $\hat{P}_f^{\mathrm{u}}$ is then written as

$$\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) = E\left[(\hat{P}_f^{\mathrm{u}} - E\left[\hat{P}_f^{\mathrm{u}}\right])^2\right]$$

$$= E\left[\left(\sum_{i=1}^{m^*-1} \mathcal{F}_i \hat{Q}(y_i) - \sum_{i=1}^{m^*-1} \mathcal{F}_i q(y_i)\right)^2\right] = E\left[\left(\sum_{i=1}^{m^*-1} \mathcal{F}_i \mathcal{E}_i\right)^2\right]$$

where we used $\hat{s}^{\mathrm{u}}(y) = \hat{q}_{y_i}$ for $y \in [y_i, y_{i+1}), i = 1, ..., m^* - 1$ and unbiasedness of the grid point estimators (Proposition 5.3.4). Reformulation yields

$$\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) = \sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \mathrm{Var}(\mathcal{E}_i) + 2 \sum_{1 \le i < j \le m^*-1} \mathcal{F}_i \mathcal{F}_j \mathrm{Cov}(\mathcal{E}_i, \mathcal{E}_j)\ . \tag{5.8}$$

Replacing the covariance according to Proposition 5.3.6 results in

$$\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) = \sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \mathrm{Var}(\mathcal{E}_i) + 2 \sum_{1 \le i < j \le m^*-1} \mathcal{F}_i \mathcal{F}_j \mathrm{Var}(\mathcal{E}_i) p_0^{j-i} \ . \tag{5.9}$$

Variance and Number of Subset Samples To proceed further, we have to analyze the variance $\mathrm{Var}(\mathcal{E}_i), i = 1, ..., m^* - 1$ where we try to rewrite it so that it is well comparable to the variance of the ordinary SuS estimator applied on the same problem formulation with same failure probability p_f. In general, by Proposition 5.3.6, $\mathrm{Var}(\mathcal{E}_i), i = 1, ..., m^* - 1$ is given by

$$\mathrm{Var}(\mathcal{E}_i) = i\sigma^2 p_0^{2(i-1)} \ . \tag{5.10}$$

The variance of SuS already looks similar, following

$$\mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) = m\sigma^2 p_0^{2(m-1)} \ . \tag{5.11}$$

However, to compare this variance to that of classical SuS, we need to account for the number of samples per subset as $\sigma^2 = \frac{p_0(1-p_0)}{N}$ (compare Assumption 5.3.1) are not necessarily the same in both equations ((5.10),(5.11)). As discussed in Remark 5.2.3, the constant number of samples per subset for a fixed total number of samples E_T generally differs in SuS and SuSI and is given by

- SuSI: $E_T = m_s N^{\mathrm{SuSI}} + 2(m^* - m_s) N^{\mathrm{SuSI}}$
- SuS: $E_T = m N^{\mathrm{SuS}}$

under re-use of seeds. Thus we have the relation

$$M^{-1} := \frac{N^{\mathrm{SuSI}}}{N^{\mathrm{SuS}}} = \frac{m}{2m^* - m_s} \ .$$

If we let $\sigma^2 = \frac{p_0(1-p_0)}{N^{\mathrm{SuS}}}$ be the variance of a subset estimator corresponding to ordinary SuS, then

$$\sigma^2(\mathrm{SuSI}) = \frac{N^{\mathrm{SuSI}}}{N^{\mathrm{SuS}}} \sigma^2 := M^{-1} \sigma^2 \ . \tag{5.12}$$

Continuation of deriving $\mathrm{Var}(\hat{P}_f^{\mathrm{u}})$ (5.9)

$$\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) = \sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \mathrm{Var}(\mathcal{E}_i) + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j \mathrm{Var}(\mathcal{E}_i) p_0^{j-i}$$

Now, using (5.10), and (5.12), neglecting the effect of changing $\bar{\gamma}_i$ for different order subsets (this effect would be beneficial for SuSI)

$$\mathrm{Var}(\mathcal{E}_i) = iM\sigma^2 p_0^{2(i-1)}$$

yields

$$\begin{aligned}\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) &= M\left(\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 i\sigma^2 p_0^{2(i-1)} + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j i\sigma^2 p_0^{2(i-1)} p_0^{j-i}\right) \\ &= M\left(\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 i\sigma^2 p_0^{2(i-1)} + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j i\sigma^2 p_0^{i+j-2}\right) .\end{aligned}$$

Now, we also know $\mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) = m\sigma^2 p_0^{2(m-1)}$ so that

$$\frac{\mathrm{Var}(\hat{P}_f^{\mathrm{u}})}{\mathrm{Var}(\hat{P}_f^{\mathrm{sus}})} = Mp_0^{-2m}\left(\sum_{i=1}^{m^*-1} \mathcal{F}_i^2 \frac{i}{m} p_0^{2i} + 2 \sum_{1\le i<j\le m^*-1} \mathcal{F}_i \mathcal{F}_j \frac{i}{m} p_0^{i+j}\right) .$$

Using Lemma 5.3.17, the right hand side can be rewritten as

$$Mp_0^{-2m} C_u \left(p_0^{2m} + 2p_0^m \mathrm{B}(\hat{P}_f^{\mathrm{u}}) + (\mathrm{B}(\hat{P}_f^{\mathrm{u}}))^2\right)$$

for some $C_u \in [\frac{1}{m}, \frac{m^*-1}{m}]$. In short, we may write

$$MC_u \left(1 + p_0^{-m} \mathrm{B}(\hat{P}_f^{\mathrm{u}})\right)^2 . \tag{5.13}$$

In addition, by Theorem 5.3.13, we have

$$\mathrm{B}(\hat{P}_f^{\mathrm{u}}) = p_f \cdot c_u \frac{1-p_0}{p_0} = p_0^m \cdot c_u \frac{1-p_0}{p_0} \tag{5.14}$$

for some $c_u \in [0,1]$. Now, combining (5.13) and (5.14) yields the claim

$$\frac{\mathrm{Var}(\hat{P}_f^{\mathrm{u}})}{\mathrm{Var}(\hat{P}_f^{\mathrm{sus}})} = MC_u \left(1 + c_u \frac{1-p_0}{p_0}\right)^2 .$$

The lower staircase case is shown accordingly.

□

Remark 5.3.19 (Variance of SuSI Under Re-use of Seeds)**.** *Under re-use of seeds, the result in Theorem 5.3.18 remains the same, except for replacing M according to Remark 5.2.3 by*

$$M = \frac{N^{SuS}}{N^{SuSI}} = \frac{m^*(2-p_0) - m_s}{m(1-p_0)} \ .$$

Theorem 5.3.18 relates the variances of SuS and SuSI under same computational costs. The different effects by bias, interpolation and differing number of samples per subset are captured in several parameters. So, to better understand the relation it is crucial to individually look at those individual parameters and see in which situations they take big or low values. This allows to identify problems where SuSI performs good or bad.

Remark 5.3.20 (Interpretation of Theorem 5.3.18)**.** *Here, we give some intuition on the theoretical findings in Theorem 5.3.18.*

- *In the upper staircase case, we have*

$$\frac{\mathrm{Var}(\hat{P}_f^{\mathrm{u}})}{\mathrm{Var}(\hat{P}_f^{\mathrm{sus}})} = \underbrace{M}_{(1)} \underbrace{C_u}_{(2)} \underbrace{\left(1 + c_u \frac{1-p_0}{p_0}\right)^2}_{(3)}$$

  for some $c_u \in [0,1]$. A factor-wise interpretation is as follows:

  (1) This is the influence of the sample number N for same total limit state evaluations E_T. The required number of subsets of SuSI$^{\mathrm{u}}$ with respect to SuS and the double evaluations necessary in SuSI subsets affect the resulting N in each subset. This again affects the Monte Carlo variance. However, we have to be aware that when re-using seeds, M is dependent on p_0 where otherwise it is not. When re-using seeds, SuSI efficiency will generally become worse in comparison to SuS for high p_0 as discussed in Remark 5.2.3.

  (2) To interpret the constant C_u, one can think of a function that is given by the product of the pdf of $\mathbf{X}_k$ and the corresponding conditional failure probability function q (as shown in Figure 5.1). Then C_u should approximately be given by the subset level m_{mass} of SuSI that locates near the 'most mass grid point' of this function divided by the total number of subset levels m of SuS, thus

$C_u \approx \frac{m_{mass}}{m}$. If the variable $\mathbf{X}_k$ is not so relevant for the failure probability, then the above importance function is almost identical to the pdf of $\mathbf{X}_k$. In this case, choosing the mean of the pdf as an approximation of the mass point is plausible. If however, the variable $\mathbf{X}_k$ has a very significant effect on the failure probability (global sensitivity index high), then the important conditional failure probabilities are taken at the tail of the pdf, yielding a low C_u as a consequence. Typically $C_u \leq 1$ and C_u can be assumed independent of p_0.

(3) This term includes the bias of the interpolation method. A higher bias generally results in higher estimated failure probabilities and thus a higher relative variation with respect to the real and fixed failure probability p_f.

Observe that the factors are not independent, as for example a high bias may result in a different number of subsets to reach a required accuracy and also change the constant c_u. While C_u decreases with respect to the relevance of $\mathbf{X}_k$, M can be decreased by proper choices of y_1 in the initial limit state and vice versa.

- *The lower staircase case yields the same components, where in contrast, the lower staircase approximation induces a negative bias and may thus result in a lower variance than an unbiased interpolation method. Accordingly, an upper boundary for the variance of the lower staircase estimation with respect to the unbiased estimator is reached for zero bias. A negative bias, resulting in a lower variance, improves the performance but also systematically underestimates the failure probability.*

Further note that the relation of c_u and c_l can not be stated in advance if f_k is free to choose.

Remark 5.3.21 (Sample Correlation Factor: Simplification in Favor of SuS)**.** *The variance of SuSI is in general even lower in comparison to SuS than stated in Theorem 5.3.18 because of the simplification by Assumption 5.3.5, where we did not account for the inner subset sample dependencies γ_i. As SuSI relies more on lower level subset estimates than SuS and higher subset levels correspond to higher dependencies γ_i, thus less effective samples, the variance of SuSI is even lower in comparison to SuS. This effect was neglected since it can depend on many factors such as convexity or concavity*

of the intermediate subset domains which are in general not observable. For less efficient implementations, such as non-adaptive MCMC, this effect can become very important as seen in Blandfort et al. (2019a).

Now that we have derived bias and variance of SuSI, we can finally analyze efficiency by the rRMSE.

Corollary 5.3.22 (rRMSE)**.** *The* rRMSE *of the failure probability estimator* $\hat{P}_f^{\mathrm{u}}$ *by SuSI is given by*

$$\mathrm{rRMSE}(\hat{P}_f^{\mathrm{u}}) = \Bigg\{ MC_u \left(1 + c_u \frac{1-p_0}{p_0}\right)^2 \frac{(1-p_0)^r}{p_0(\log(p_0))^2} \frac{(\log(p_f))^2}{E_T}(1+\bar{\gamma}) + \left(c_u \frac{1-p_0}{p_0}\right)^2 \Bigg\}^{\frac{1}{2}}$$

with C_u, c_u *as in Theorem 5.3.18. If seeds are not re-used,* $r = 1$ *and* M *as in Theorem 5.3.18. Otherwise, we have* $r = 2$ *and* M *as in Remark 5.3.19. The factor* $(1+\bar{\gamma})$ *accounts for the effective sample number as in Theorem 4.1.11.* C_u *and* c_u *can be replaced by* $C_l, -c_l$ *respectively for the lower staircase case* $\hat{P}_f^{\mathrm{l}}$*.*

Proof. We have

$$\mathrm{rRMSE}(\hat{P}_f^{\mathrm{u}}) = \sqrt{\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) + (\mathrm{B}(\hat{P}_f^{\mathrm{u}}))^2}\ .$$

By Theorem 5.3.13 and Theorem 5.3.18 this becomes

$$\frac{1}{p_f} \cdot \sqrt{MC_u \left(1 + c_u \frac{1-p_0}{p_0}\right)^2 \mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) + \left(c_u \frac{1-p_0}{p_0} p_f\right)^2}\ .$$

Using Theorem 4.1.11

$$\mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) = \Big(\mathrm{CV}(\hat{P}_f, E_T)\Big)^2 p_f^2 = \frac{(1-p_0)^r}{p_0(\log(p_0))^2} \frac{(\log(p_f))^2}{E_T}(1+\bar{\gamma}) p_f^2$$

then yields the result. □

It is good to have an explicit formula for the rRMSE in Corollary 5.3.22. However, we are more interested in the relation of rRMSE of SuSI and variance of SuS under optimal parameter settings. We already derived the relation of the variances of SuSI and SuS explicitly. So, we could either additionally compare the bias of SuSI with the variance of SuS or just use the

explicit formula for the rRMSE of SuSI in Corollary 5.3.22 and compare its results to that of the variance of SuS in Theorem 4.1.11 directly. We choose the latter approach, where we will also consider the optimal p_0 for both approaches separately so that each method reaches its best performance on a specific given problem. The choice of an optimal p_0 in SuSI obviously depends on the explicit case, e.g. the total number of evaluations E_T, because if the variance becomes low, the bias has a higher effect on the absolute value of the errors and therefore choosing a higher p_0 to decrease the bias becomes more important for a good performance of SuSI. The statistical derivations in the staircase approach, provide a general result for the rRMSE dependent on parameters of the problem at hand. It remains to consider results for specific scenarios and compare the rRMSE of SuS and SuSI in realistic situations. This will allow to provide more insights on practical efficiency of the new algorithm. Conclusions are jointly drawn for staircase and interpolation by splines approaches in Section 5.4. Furthermore an empirical analysis by a simulation study considering artificial and real world problems is given in Section 6.1.

5.3.3 Piecewise Cubic Hermite Interpolation

Now that we have examined the staircase interpolation approach, which yields boundaries with respect to systematic estimation errors of the failure probability, we next study a more sophisticated interpolation method, which should give more accurate results in many cases. Interpolation yields smooth and visually pleasing functions for approximation. We thus achieve good accuracy by interpolation, if either the grid point distances are small or the interpolated function is smooth.

Remark 5.3.23 (Smoothness of q)**.** *According to the construction of q under the influence of many random variables, it is often plausible that q is rather smooth. The idea is that the noise of the stochastic variables flattens abrupt changes in the underlying function. We assume q to be continuously differentiable up to at least fifth order on the domain of $\mathbf{X}_k$, $q \in C^j(D_k), j \geq 5$. Although violations of this assumption may prevent us from deriving theoretical conclusions at some points, practically it will still be beneficial to use the interpolation approach in many such cases. However, we need to take a heuristic approach here and do not claim any general validity here.*

We thus study piecewise cubic Hermite interpolation as an interpolation method in SuSI. In our case, we however certainly know that q is monotone. Thus, we should also examine shape preserving interpolation methods, where we want to keep a locality property of the approximation method for easier analysis. The proposed method will be the one validated in our simulation studies and is therefore motivated by also providing some historical background, demonstrating it is well tested and easy to integrate in implementation. Furthermore, this simplifies further study of the method for the interested reader. Fortunately, monotonicity preserving interpolation was examined by Fritsch and Carlson (1980), decades ago. It was shown, how one can, only by modifying the interpolation derivatives at the grid points, get a monotone piecewise cubic Hermite interpolation. The result was extended in Fritsch and Butland (1984) by using results obtained in Butland (1980), offering a simple formula for determination of the derivatives which fulfilled all conditions given in Fritsch and Carlson (1980) and was also shown to lead to 'visually pleasing' results. This procedure is not only simple and can be applied for all monotone data, it is also completely local and does account for different scales, which we have in Subset Simulation Interpolation, very well. The method for determination of the derivatives is also known as the

weighted Harmonic mean:

$$\frac{dq(y_i)}{dy} = \frac{d_{i-1}d_i}{\alpha d_i + (1-\alpha)d_{i-1}}$$

for $d_{i-1}, d_i > 0$ and 0, otherwise. The

$$d_i = \frac{q(y_{i+1}) - q(y_i)}{y_{i+1} - y_i}$$

are the finite differences and the weights are given as

$$\alpha = \frac{y_i - y_{i-1} + 2(y_{i+1} - y_i)}{3(y_{i+1} - y_{i-1})} \ .$$

For equidistant points, this becomes just the Harmonic mean. While the Harmonic mean itself, although being not so popular in general, was already known (see e.g. Coggeshall (1886), Ferger (1931)), its perfect match to this application was a very nice and pleasant discovery. Although this method was explored decades ago, it is still a good choice for monotone interpolation and is also frequently used in current software implementations (see e.g. Python; in scipy:'PchipInterpolator' or Matlab:'pchip', entitled as Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)). By using this method, we therefore benefit, beyond the shape preserving property, from the fact that it is well explored and also that it is already implemented in several software packages, making it easy applicable for our new method without additional effort. Note, the guarantee of monotonicity avoids overshooting.

Remark 5.3.24 (A Potential Drawback)**.** *Constructing derivatives by Harmonic means tends to create too flat curves (Fritsch and Butland (1984)). In addition, the interpolation function is only C^1 (Fritsch and Butland (1984)) instead of C^2 as in usual cubic spline interpolation, since the splines need to fulfill the monotonicity condition and are thus not chosen as smooth as they could otherwise. However, it has many benefits such as $O(h^2)$ convergence for uniform data, 'visually pleasing' results and being simple and completely local. Still, the tendency to create too flat curves should be noted because this may result in a systematic underestimation of the failure probability, a negative bias. A more detailed discussion for accurateness of monotone cubic interpolation is given in Huynh (1993).*

In this section, we analyze the properties of SuSI with respect to shape preserving and standard piecewise cubic Hermite interpolation. These methods allow for derivation of good results when the conditional failure function q is smooth. But the method can not guarantee a positive bias such as the upper staircase approximation. Thus, in contrast to the staircase approach, even a systematic underestimation of the failure probability can be the consequence. On the other hand, we expect more accurate results in many cases. Also Subset Simulation has a potential for underestimation of the failure probability as discussed in Section 3.2. The interpolation methods in this section are generally supposed to offer good performance on average, but usually lack some control of the bias. Other methods should be used additionally.

Remark 5.3.25. *Derivation of the results in this section is, in contrast to the staircase approach, often based on numerical methods. We will often need to look at the permitted values for p_0 and focus on intervals that are typically more important for evaluation of small failure probabilities. Special features at the domain boundaries will not be considered as the corresponding effects will usually be negligible.*

In the proposed interpolation method of this section, each interval $[y_w, y_{w+1}) \subset D_k$, corresponds to a piecewise polynomial as given in the following definition.

Definition 5.3.26 (Hermite Interpolation)**.** *For $w = 1, ..., m^* - 1$, the cubic Hermite interpolation splines $s_w^{\text{Spl.}}(y)$, $y \in I_w := [y_w, y_{w+1}) \subset D_k$, are given by*

$$s_w^{\text{Spl.}}(y) = q_w H_{w,1}(y) + q_{w+1} H_{w,2}(y) + m_w H_{w,3}(y) + m_{w+1} H_{w,4}(y) \quad (5.15)$$

with $q_w = q(y_w)$, m_w the tangent at grid point y_w and

$$H_{w,1}(y) = 3\left(\frac{y_{w+1} - y}{y_{w+1} - y_w}\right)^2 - 2\left(\frac{y_{w+1} - y}{y_{w+1} - y_w}\right)^3,$$

$$H_{w,2}(y) = 3\left(\frac{y - y_w}{y_{w+1} - y_w}\right)^2 - 2\left(\frac{y - y_w}{y_{w+1} - y_w}\right)^3,$$

$$H_{w,3}(y) = -(y_{w+1} - y_w)\left(\left(\frac{y_{w+1} - y}{y_{w+1} - y_w}\right)^3 - \left(\frac{y_{w+1} - y}{y_{w+1} - y_w}\right)^2\right) \text{ and}$$

$$H_{w,4}(y) = (y_{w+1} - y_w)\left(\left(\frac{y - y_w}{y_{w+1} - y_w}\right)^3 - \left(\frac{y - y_w}{y_{w+1} - y_w}\right)^2\right).$$

By Definition 5.3.26, the function value is thus given as a linear combination of the Hermite basis functions $H_{w,1}, H_{w,2}, H_{w,3}, H_{w,4}$ for every $w = 1, ..., m^* - 1$.

Assumption 5.3.27. *The grid points y_j, $j = 1, ..., m^*$ are assumed to be equidistant*

$$y_{w+1} - y_w = h \text{ or equivalently } |I_w| = h$$

for all $w = 1, ..., m^* - 1$ and some $h \in \mathbb{R}^+$. I_w as in Definition 5.3.26.

The intervals for upper and lower part of the domain $I_0 \subseteq (-\infty, y_1)$ and $I_{m^*} \subseteq (y_{m^*}, \infty)$ are neglected. This is admissible according to Lemma 5.3.9 for appropriate choices of y_1 and y_{m^*}.

Using equidistant points, the cubic Hermite interpolation splines in Definition 5.3.26 become

$$\begin{aligned}
H_{w,1}(y) &= 3\left(\frac{y_{w+1} - y}{h}\right)^2 - 2\left(\frac{y_{w+1} - y}{h}\right)^3, \\
H_{w,2}(y) &= 3\left(\frac{y - y_w}{h}\right)^2 - 2\left(\frac{y - y_w}{h}\right)^3, \\
H_{w,3}(y) &= -h\left(\left(\frac{y_{w+1} - y}{h}\right)^3 - \left(\frac{y_{w+1} - y}{h}\right)^2\right) \text{ and} \\
H_{w,4}(y) &= h\left(\left(\frac{y - y_w}{h}\right)^3 - \left(\frac{y - y_w}{h}\right)^2\right).
\end{aligned}$$

The key for a well suitable cubic Hermite interpolation is the determination of the tangents $m_w, w = 1, ..., m^*$. We consider three different types for tangent derivation:

(3FD) Three-point central finite difference. This approach is the classical one for cubic Hermite spline interpolation and is easy to analyze due to its simple representation. Here, tangents are given by slope

$$m_w^{\text{3FD}} = \frac{1}{2}\left(\frac{q_{w+1} - q_w}{y_{w+1} - y_w} + \frac{q_w - q_{w-1}}{y_w - y_{w-1}}\right) = \frac{q_{w+1} - q_{w-1}}{2h}.$$

(5FD) Five-point central finite difference:

$$m_w^{\text{5FD}} = \frac{q_{w-2} - 8q_{w-1} + 8q_{w+1} + q_{w+2}}{12h}$$

This approach allows for increased accuracy for smooth functions in comparison to 3FD at the cost of an increased complexity in statistical analysis.

(HM) Harmonic mean of the finite differences. We define

$$m_w^{\mathrm{HM}} = \frac{2d_{w-1}d_w}{d_{w-1} + d_w} \tag{5.16}$$

with

$$d_w = \frac{q_{w+1} - q_w}{y_{w+1} - y_w} = \frac{q_{w+1} - q_w}{h}$$

the two point finite differences. In comparison to the other approaches, this one is generally harder to handle in the analysis due to the fraction term but on the other hand it preserves monotonicity, which suits our demands very well.

The formulas for the tangent slopes can not be applied on all grid points, instead they only e.g. hold for $w = 2, ..., m^* - 1$ for (3FD). At the end points we need another definition such as $\frac{q_{w+1}-q_w}{h}$ for m_w. This is neglected here as relevance of the corresponding intervals should be small anyways, in order to have accurate results.

In order to analyze the accuracy of SuSI, we consequently need to account for the accuracy of the chosen method for tangent approximation. Those can be derived by considering the corresponding Taylor approximations and are particularly interesting since they will appear in the bias term later.

Remark 5.3.28 (Decomposition of the Estimation Error)**.** *In Subset Simulation Interpolation with cubic Hermite spline interpolation, we are required to choose a specific method (3FD, 5FD, HM) for derivation of the derivatives at the grid points. Then the error made by approximating the true conditional failure probability $q(y)$ by the estimation, e.g. choosing method 3FD resulting in $\hat{S}_w^{\mathrm{Spl.3FD}}(y)$ can be split in a stochastic error and two deterministic errors. The deterministic errors are given by the error that results from assuming a non-correct derivative as consequence of applying method 3FD, 5FD or HM and the error that is present by generally applying cubic Hermite spline*

interpolation for approximation of q. We can thus specify the error as follows

$$\hat{S}_w^{\text{Spl.3FD}}(y) - q(y) = \underbrace{\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y)}_{stoch.} + \underbrace{s_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.}}(y)}_{deriv.} + \underbrace{s_w^{\text{Spl.}}(y) - q(y)}_{interp.} ,$$

so that we may separately analyze the specific errors. In the following analysis, we consider separately stochastic error

$$\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y)$$

and deterministic error (bias)

$$\text{B}(\hat{P}_f^{\text{Spl.3FD}}) = s_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.}}(y) + s_w^{\text{Spl.}}(y) - q(y)$$

by derivative approximation as well as general interpolation by cubic Hermite spline interpolation. The sum of derivative error and interpolation error yields the bias at y, having

$$\text{B}(\hat{P}_f^{\text{Spl.3FD}}) = \sum_{i=1}^{m^*-1} \int_{y_w}^{y_{w+1}} f_k(y)(s_w^{\text{Spl.3FD}}(y) - q(y))dy$$

for the bias of the SuSI estimation $\hat{P}_f^{\text{Spl.3FD}}$.

The individual errors in Remark 5.3.28 are first separately examined, later putting together the complete result by all single component results.

Definition 5.3.29. *We write $q^{(j)}$ for the j-th derivative of the conditional failure probability function q,*

First, we examine the derivative error for the most simple approximation.

Lemma 5.3.30 (Accuracy of Deterministic Tangent Determination)**.** *The tangent determinations of method (3FD) and method (5FD) have the following accuracies.*

3FD: Approach 3FD yields

$$m_w^{3FD} = q_w^{(1)} + \frac{h^2}{3} q^{(3)}(\tau_{w,3FD})$$

for a $\tau_{w,3FD} \in [y_{w-1}, y_{w+1}]$.

5FD: Using 5FD for approximation results in

$$m_w^{5FD} = q_w^{(1)} + \frac{h^4}{180}\Big(\big(q^{(5)}(\tau_1) + q^{(5)}(\tau_2)\big) - 4\,\big(q^{(5)}(\tau_3) + q^{(5)}(\tau_4)\big) \Big)$$

with $\tau_1 \in [y_w, y_{w+1}]$, $\tau_2 \in [y_{w-1}, y_w]$, $\tau_3 \in [y_w, y_{w+2}]$ *and* $\tau_4 \in [y_{w-2}, y_w]$ *or under a change of representation in*

$$m_w^{5FD} \in \left[q_w^{(1)} - \Big|\frac{h^4}{36} q^{(5)}(\tau_{w,5FD})\Big|, q_w^{(1)} + \Big|\frac{h^4}{36} q^{(5)}(\tau_{w,5FD})\Big| \right]$$

for a $\tau_{w,5FD} \in [y_{w-2}, y_{w+2}]$.

Proof. A proof is given in Appendix C. □

So we have $O(h^2)$ convergence of 3FD and $O(h^4)$ convergence of 5FD to the true value. By that, increasing p_0 which results in decreasing h will allow convergence to the real derivative at the grid points for interpolation. However, choosing high p_0 to decrease h will increase the computational cost drastically in most cases. The real benefit of course also depends on the relation of p_0 and h and might induce expensive reductions of the interval length. As a consequence in the case of one run of SuSI, it will generally be much more important to have smooth conditional failure probability functions q to receive a small error. In more detail, we have an error term that is given by a fraction of a third derivative in case of 3FD, or even of fifth derivatives if we use the more sophisticated approach 5FD.

Lemma 5.3.31. *Under the assumption of equidistant points* $y_{w+1} - y_w = h$ *for all* $w = 1, ..., m^* - 1$, $q_w^{(1)} \neq 0$ *and small* h, *approximation of* m_w *by (HM)*

yields

$$m_w^{HM} = q_w^{(1)} + \Bigg(h^2 \underbrace{\left[q_w^{(1)} \left(q^{(3)}(\tau_3) + q^{(3)}(\tau_4)\right) - 3(q_w^{(2)})^2\right]}_{r_2} + h^3 \underbrace{\left[q_w^{(2)} \left(q^{(3)}(\tau_4) - q^{(3)}(\tau_3)\right)\right]}_{r_3} + h^4 \underbrace{\left[\frac{1}{3} q^{(3)}(\tau_3) q^{(3)}(\tau_4)\right]}_{r_4} \Bigg) \cdot \underbrace{\frac{1}{12 q_w^{(1)} + h^2 \left(q^{(3)}(\tau_3) + q^{(3)}(\tau_4)\right)}}_{r^*} .$$

Proof. Appendix C. □

By Lemma 5.3.31, we have $O(h^2)$ convergence with respect to the grid point distances but might also consider smoothness of q to argue on sufficient well approximations, due to dependence of the error term on derivatives of higher orders only. We have $r_i \cdot r^*$ is $O(h^i)$ for $i = 2, 3, 4$ if $q_w^{(1)} \neq 0$ and $O(h^{i-2})$ otherwise, where the latter certainly holds if q is strictly monotone. Although this might appear to be a worse deterministic error than for 3FD, the deterministic error is not tractable anyways and HM remains a favorable approach since it utilizes the given information about the real shape of q, providing monotone functions as a result only.

Remark 5.3.32. *In the following, we will use the big $O-$notation for the bias terms, as the focus of analysis is on the stochastic part. The deterministic error by interpolation is important, but since the shape of the underlying real function q is generally unknown and particularly no information about smoothness is available, one should not give results on the systematic errors too much weight in the conclusions. As a consequence, we will specify the bias by terms such as e.g. $O(q^{(3)}h^2)$ for accuracy up to terms containing derivatives of third order multiplied by the squared grid point distance.*

Next, we analyze the statistical properties of SuSI when interpolating with one of the methods and compare their efficiency to that of ordinary SuS.

Proposition 5.3.33 (Hermite Representation, Estimation Error (3FD))**.** *The estimation error by cubic Hermite spline interpolation with method 3FD, with equidistant grid points, i.e.* $h = y_w - y_{w-1}$*, is given by*

$$\begin{aligned}\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.}}(y) =& \mathcal{E}_w H_{w,1}(y) + \mathcal{E}_{w+1} H_{w,2}(y) + \frac{\mathcal{E}_{w+1} - \mathcal{E}_{w-1}}{2h} H_{w,3}(y) \\ &+ \frac{\mathcal{E}_{w+2} - \mathcal{E}_w}{2h} H_{w,4}(y) + O(h^2 \cdot q^{(3)})\end{aligned}$$

for $y \in [y_w, y_{w+1}]$*,* $s_w^{\text{Spl.}}(y)$ *as in* (5.15) *and* $\hat{S}_w^{\text{Spl.3FD}}(y)$ *being the corresponding estimator. We also refer to the estimation error* $\frac{\mathcal{E}_{w+1} - \mathcal{E}_{w-1}}{2h}$ *as* $\hat{m}_{w,err}^{3FD}$ *and define* $\hat{m}_{w+1,err}^{3FD}$ *correspondingly.* $\mathcal{E}_w = \hat{Q}(y_w) - q(y_w)$ *is the estimation error of the failure probability at grid point* y_w*.*

Proof. First, we replace the general definition of the spline (5.15) with its estimation

$$\hat{S}_w^{\text{Spl.3FD}}(y) = \hat{Q}_w H_{w,1}(y) + \hat{Q}_{w+1} H_{w,2}(y) + \hat{m}_w H_{w,3}(y) + \hat{m}_{w+1} H_{w,4}(y) \ . \tag{5.17}$$

Now by Proposition 5.3.4, the estimator can be rewritten by real value and Gaussian noise as

$$\hat{Q}_w = q_w + \mathcal{E}_w \tag{5.18}$$

where $\mathcal{E}_w \sim \mathcal{N}(0, \sigma_{\mathcal{E}_w}^2)$ with variance $\sigma_{\mathcal{E}_w}^2$ as in Proposition 5.3.6. For the derivatives, additionally using Lemma 5.3.30, we thus get

$$\hat{m}_w = \frac{q_{w+1} + \mathcal{E}_{w+1} - (q_{w-1} + \mathcal{E}_{w-1})}{2h} + O(h^2 \cdot q^{(3)}) \ . \tag{5.19}$$

Plugging (5.18) and (5.19) into (5.17) then directly yields the result together with (5.15). □

Proposition 5.3.34 (Hermite Representation, Estimation Error (HM))**.** *The estimation error by cubic Hermite spline interpolation with method HM is given by*

$$\begin{aligned}\hat{S}_w^{\text{Spl.HM}}(y) - s_w^{\text{Spl.}}(y) =& \mathcal{E}_w H_{w,1}(y) + \mathcal{E}_{w+1} H_{w,2}(y) + \hat{m}_{w,err}^{HM} H_{w,3}(y) \\ &+ \hat{m}_{w+1,err}^{HM} H_{w,4}(y) + O(h^2 \cdot q^{(2)})\end{aligned}$$

for $y \in I_w$, $s_w^{\text{Spl.}}(y)$ as in (5.15) and $\hat{S}_w^{\text{Spl.HM}}(y)$ the corresponding estimator. $\hat{m}_{w,err}^{HM}$ is given by

$$\frac{2\left(-p_0^2\mathcal{E}_{w-1} + \mathcal{E}_w(p_0^2-1) + \mathcal{E}_{w+1}\right)}{h(p_0+1)^2}$$

and $\hat{m}_{w+1,err}^{HM}$ analogously.

Proof. Appendix C. □

In the first place Proposition 5.3.33 and Proposition 5.3.34, i.e. the estimation errors in 3FD and HM, are not so easy to compare. However, differences are the estimated derivative values $\hat{m}_{w,err}^{3FD}$ and $\hat{m}_{w,err}^{HM}$ as well as having accuracy up to $O(h^2q^{(3)})$ terms for 3FD and up to $O(h^2q^{(2)})$ terms in case of approximation by HM. So we need to compare the estimated derivative values where we can include our knowledge about the estimation errors $\mathcal{E}_{w-1}, \mathcal{E}_w$ and $\mathcal{E}_{w+1}$ in analysis next. The estimation errors are given as

$$\mathcal{E}_w = \sum_{i=1}^{w} \sigma_i \prod_{j\neq i} p_0 Z_i = \sum_{i=1}^{w} \sigma_i p_0^{w-1} Z_i \tag{5.20}$$

and $\mathcal{E}_{w-1}, \mathcal{E}_{w+1}$ analogously, compare Proposition 5.3.4. The resulting relation of the derivative variances for both approaches is derived in Appendix C and illustrated in Figure 5.7. Numerical evaluation yields similar variances for high $p_0 \geq 0.5$ and significantly lower variances for low p_0. The result reacts insensitively to changes in $w > 1$ which covers all relevant constellations most of the time. We are now able to draw a conclusion on the comparison of the two approaches for our purpose.

Remark 5.3.35 (Comparison of Approaches by 3FD and HM)**.** *The estimation errors by cubic Hermite spline interpolation with method 3FD and HM (compare Proposition 5.3.33 and Proposition 5.3.34) should generally have similar variances. For low intermediate subset probabilities p_0, HM might even outperform 3FD. Keep in mind that only half of the main terms in Hermite representation differ and the methods are similar in general. Thus, the variances should also be comparable and not differ as much as indicated by Figure 5.7. The same holds for the bias, which also does only affect some error terms in interpolation.*

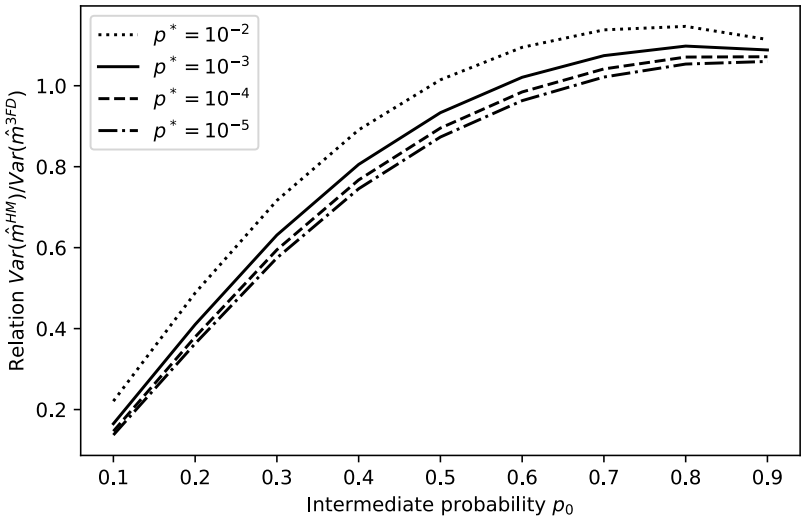


Figure 5.7: Comparison of the relation of $\mathrm{Var}(\hat{m}^{HM}_{w,err})$ and $\mathrm{Var}(\hat{m}^{3FD}_{w,err})$ with respect to different choices of p_0. Additionally, several w were considered by $w = \frac{\log(p^*)}{\log(p_0)}$ so that comparability of different p_0 is given. The artificially introduced p^* then defines the probability where estimation is performed and only serves for defining w.

Concluding, we see from Remark 5.3.35 that the stochastic error by HM should in general be lower than that by 3FD estimation as we have the same characteristics except for the higher variance in case of 3FD at the derivative estimation errors $\mathrm{Var}(\hat{m}_{k,err}^{HM}) < \mathrm{Var}(\hat{m}_{k,err}^{3FD})$. This inequality is not generally fulfilled, but holds for typical settings concerning p_0 and the most relevant estimation intervals of the interpolation.

As a consequence, we restrict ourselves to the case of choosing the 3FD approach, remembering that HM will not be worse in general and most likely take similar results. We are particularly interested in general variation of the splines in comparison to SuS, so that small differences need not be accounted for in detail. The following part of the section is dedicated to the analysis of the statistical properties of the 3FD method, with the idea that the derived results also approximately apply to the other approaches as well. To proceed, we first simplify notation.

Definition 5.3.36 (Interval Estimation Error)**.** *Define the estimation error for cubic Hermite spline interpolation with respect to the integral between two successive grid points y_w, y_{w+1} as*

$$\mathcal{S}_w := \int_{y_w}^{y_{w+1}} f_k(y)(\hat{S}_w^{\mathrm{Spl.}}(y) - s_w^{\mathrm{Spl.}}(y))dy$$

with $s_w^{\mathrm{Spl.}}(y)$ the cubic Hermite interpolation spline on interval $[y_w, y_{w+1}]$ and $\hat{S}_w^{\mathrm{Spl.}}(y)$ its corresponding estimator such as $\hat{S}_w^{\mathrm{Spl.3FD}}(y)$, $\hat{S}_w^{\mathrm{Spl.5FD}}(y)$ or $\hat{S}_w^{\mathrm{Spl.HM}}(y)$.

Proposition 5.3.37. *Under Assumption 5.3.5, the interval estimation error by cubic Hermite spline interpolation (3FD) has zero mean and variance*

$$L_w = \left(\int_{y_w}^{y_{w+1}} (f_k(y))^2 dy\right) h\mathrm{Var}(\mathcal{E}_w) f_1(p_0, w)$$

with

$$\lim_{w\to\infty} f_1(p_0, w) \leq f_1(p_0, w) \leq f_1(p_0, 1) \ ,$$

where

$$f_1(p_0, 1) = \frac{1}{210}\left(\frac{3}{2}p_0^4 - \frac{47}{2}p_0^3 + \frac{325}{2}p_0^2 + 76p_0 + 85\right)$$

and

$$\lim_{w\to\infty} f_1(p_0, w) = \frac{1}{210}\left(\frac{1}{2}p_0^4 - \frac{47}{4}p_0^3 + \frac{155}{2}p_0^2 + \frac{307}{4}p_0 - \frac{47}{4}p_0^{-1} + \frac{1}{2}p_0^{-2} + \frac{155}{2}\right).$$

As a short notation, we use $f_1(p_0, \infty) := \lim_{w\to\infty} f_1(p_0, w)$.

Proof. We start with the Hermite Representation, given in Proposition 5.3.33 as

$$\begin{aligned}\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y) =& \mathcal{E}_w H_{w,1}(y) + \mathcal{E}_{w+1} H_{w,2}(y) + \frac{\mathcal{E}_{w+1} - \mathcal{E}_{w-1}}{2h} H_{w,3}(y) \\ &+ \frac{\mathcal{E}_{w+2} - \mathcal{E}_w}{2h} H_{w,4}(y) \ .\end{aligned}$$

This is a sum of random variables with zero mean, so that the mean of the integrands is also zero and their variance is given by the sum of the covariances of the corresponding random variables. This also holds for the whole integral, so that

$$\mathcal{S}_w = \int_{y_w}^{y_{w+1}} f_k(y)(\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y))dy$$

has also zero mean for every $w = 1, 2, ..., m^* - 1$. The variance, defined by $L_w \in \mathbb{R}^+$, will be evaluated next.

To do so, note that by definition of the Hermite Basis functions, we have for $\varphi(t) := 3t^2 - 2t^3$, $\psi(t) := t^3 - t^2$ and $t = \frac{y_{w+1}-y}{h}$:

$$\begin{aligned}H_{w,1}(y) &= \varphi(t) \\ H_{w,2}(y) &= \varphi(1-t) \\ H_{w,3}(y) &= -h\psi(t) \\ H_{w,4}(y) &= h\psi(1-t) \ .\end{aligned}$$

This yields

$$\begin{aligned}\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y) = & \underbrace{\mathcal{E}_w}_{=:e_{w,1}} \varphi(t) + \underbrace{\mathcal{E}_{w+1}}_{=:e_{w,2}} \varphi(1-t) \\ & \underbrace{-\frac{1}{2}(\mathcal{E}_{w+1} - \mathcal{E}_{w-1})}_{=:e_{w,3}} \psi(t) \\ & + \underbrace{\frac{1}{2}(\mathcal{E}_{w+2} - \mathcal{E}_w)}_{=:e_{w,4}} \psi(1-t) \\ & + O(h^2 \cdot q^{(3)})\end{aligned} \tag{5.21}$$

for $y \in [y_w, y_{w+1}]$ and corresponding $t \in [0,1]$. Since we are interested in analyzing the variance of $\mathcal{S}_w$ given by

$$\mathrm{Var}(\mathcal{S}_w) = E\left[\left(\int_{y_w}^{y_{w+1}} f_k(y)(\hat{S}_w^{\mathrm{Spl.3FD}}(y) - s_w^{\mathrm{Spl.3FD}}(y))\right)^2\right] ,$$

we use Cauchy Schwarz for simplification

$$\begin{aligned}\mathrm{Var}(\mathcal{S}_w) &\leq E\left[\left(\int_{y_w}^{y_{w+1}} (f_k(y))^2 dy\right)\left(\int_{y_w}^{y_{w+1}} (\hat{S}_w^{\mathrm{Spl.3FD}}(y) - s_w^{\mathrm{Spl.3FD}}(y))^2 dy\right)\right] \\ &= \left(\int_{y_w}^{y_{w+1}} (f_k(y))^2 dy\right) E\left[\left(\int_{y_w}^{y_{w+1}} (\hat{S}_w^{\mathrm{Spl.3FD}}(y) - s_w^{\mathrm{Spl.3FD}}(y))^2 dy\right)\right] .\end{aligned}$$

So, it remains to analyze the term under the expectation to show the claim. By substituting $t = \frac{y_{w+1}-y}{h}$ we have by (5.21)

$$\begin{aligned}&E\left[\left(\int_{y_w}^{y_{w+1}} (\hat{S}_w^{\mathrm{Spl.3FD}}(y) - s_w^{\mathrm{Spl.3FD}}(y))^2 dy\right)\right] \\ &\quad = h \cdot E\left[\int_0^1 \Big(e_{w,1}\varphi(t) + e_{w,2}\varphi(1-t) \right. \\ &\qquad\qquad \left. + e_{w,3}\psi(t) + e_{w,4}\psi(1-t)\Big)^2 dt\right] ,\end{aligned}$$

where before taking the expectation, we need to evaluate the integral. This is a straightforward but lengthy calculation, where one just needs to expand the product, use linearity of the integral and then calculate the result for the individual polynomial integrals, keeping the pre-factors. The evaluation yields

$$\begin{aligned}&h \cdot E\left[\frac{2e_{w,4}^2 + (3e_{w,3} - 22e_{w,2} - 13e_{w,1})e_{w,4} + 2e_{w,3}^2}{210}\right. \\ &\left.\frac{-(13e_{w,2} + 22e_{w,1})e_{w,3} + 78e_{w,2}^2 + 54e_{w,1}e_{w,2} + 78e_{w,1}^2}{210}\right]\end{aligned} \tag{5.22}$$

with

$$\begin{aligned} e_{w,1} =& \mathcal{E}_w \ , \\ e_{w,2} =& \mathcal{E}_{w+1} \ , \\ e_{w,3} =& -\frac{1}{2}(\mathcal{E}_{w+1} - \mathcal{E}_{w-1}) \ , \\ e_{w,4} =& \frac{1}{2}(\mathcal{E}_{w+2} - \mathcal{E}_w) \ . \end{aligned}$$

Linearity of the expected value then allows to derive the result by Proposition 5.3.6, resulting in

$$\begin{aligned} =& h\frac{\mathrm{Var}(\mathcal{E}_w)}{210} \cdot \left(\frac{1}{2}\frac{w+2}{w}p_0^4 - \frac{47}{4}\frac{w+1}{w}p_0^3 + 85\frac{w+1}{w}p_0^2 - \frac{15}{2}p_0^2 + \frac{3}{4}\frac{w-1}{w}p_0 \right. \\ & \left. + 76p_0 - \frac{47}{4}\frac{w-1}{w}p_0^{-1} + \frac{1}{2}\frac{w-1}{w}p_0^{-2} - \frac{15}{2}\frac{w-1}{w} + 85 \right) . \end{aligned} \tag{5.23}$$

For more details on derivation of (5.22) and (5.23), see Appendix C. Define $f_1(p_0, w) := (5.23) \cdot \frac{1}{h\mathrm{Var}(\mathcal{E}_w)}$ the term inside the brackets multiplied by $\frac{1}{210}$. We can not draw conclusions independent of w here, but we can derive approximate values or boundaries instead. Its derivative with respect to w is given by

$$\frac{df_1(p_0, w)}{dw} = \frac{-4p_0^4 + 47p_0^3 - 340p_0^2 + 3p_0 - 47p_0^{-1} + 2p_0^{-2}}{840w^2} \ .$$

It is easy to check numerically that $\frac{df_1(p_0,w)}{dw} < 0$ for all $p_0 \geq 0.0425$ and positive w so that for relevant values of p_0 and w, $f_1(p_0, w)$ is monotone decreasing with respect to w. As a consequence, for $0.0425 \leq p_0 < 1$ we can bound $f_1(p_0, w)$ by

$$\lim_{w\to\infty} f_1(p_0, w) \leq f_1(p_0, w) \leq f_1(p_0, 1)$$

which yields the claim. □

To better understand the conclusion in Proposition 5.3.37, Figure 5.8 shows typical values and boundaries for function values of $f_1(p_0, w)$ with respect to w and p_0.

Another relevant integral is examined in the following proposition.

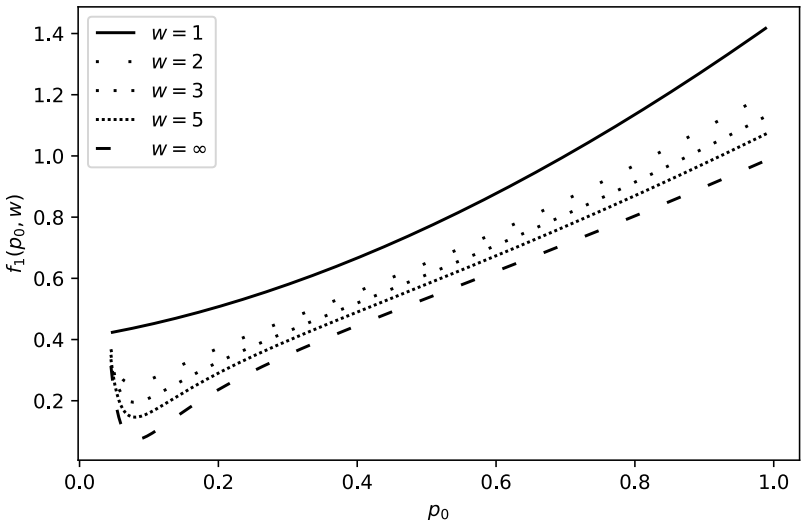


Figure 5.8: Illustration of the boundaries of $f_1(p_0, w)$ in Proposition 5.3.37. The lower boundary corresponds to a high subset level w and the upper boundary to small w.

Proposition 5.3.38. *Under Assumption 5.3.5, the cubic Hermite spline interpolation value for integration of the interval between two grid points y_w and $y_{w+1}, w = 1, ..., m^* - 1$ is given by*

$$\int_{y_w}^{y_{w+1}} (s_w^{\text{Spl.3FD}}(y))^2 dy = h \cdot p_0^{2w} f_1(p_0, \infty)$$

with $f_1(p_0, w)$ as in Proposition 5.3.37.

Proof. The proof is similar to Proposition 5.3.37, but more simple since we do not need to evaluate stochastic terms. First, in Hermite representation we have

$$s_w^{\text{Spl.3FD}}(y) = q_w H_{w,1}(y) + q_{w+1} H_{w,2}(y) + m_w^{3FD} H_{w,3}(y) + m_{w+1}^{3FD} H_{w,4}(y) \ .$$

Replacing m_w^{3FD}, m_{w+1}^{3FD} and assuming equidistant grid points results in

$$\begin{aligned} s_w^{\text{Spl.3FD}}(y) =& q_w H_{w,1}(y) + q_{w+1} H_{w,2}(y) + \frac{q_{w+1} - q_{w-1}}{2h} H_{w,3}(y) \\ &+ \frac{q_{w+2} - q_w}{2h} H_{w,4}(y) \ . \end{aligned}$$

Continuing with

$$\int_{y_w}^{y_{w+1}} (s_w^{\text{Spl.3FD}}(y))^2 dy$$

we can rewrite with substitution of $t = \frac{y_{w+1}-y}{h}$ and $\varphi(t) := 3t^2 - 2t^3$, $\psi(t) := t^3 - t^2$ as above to receive

$$h \cdot \int_0^1 \Bigg(q_w \varphi(t) + q_{w+1}\varphi(1-t) - \frac{1}{2}(q_{w+1} - q_{w-1})\psi(t) + \frac{1}{2}(q_{w+2} - q_w)\psi(1-t) \Bigg)^2 dt$$

which might again be evaluated as above, only differing in having the conditional failure probability at the grid points instead of the corresponding error terms. The result is given by

$$\frac{h}{210}\Big[2(\frac{1}{2}(q_{w+2} - q_w))^2 + (3\frac{1}{2}(q_{w+1} - q_{w-1}) - 22q_{w+1} - 13q_w)\frac{1}{2}(q_{w+2} - q_w) + 2(\frac{1}{2}(q_{w+1} - q_{w-1}))^2 - (13q_{w+1} + 22q_w)\frac{1}{2}(q_{w+1} - q_{w-1}) + 78q_{w+1}^2 + 54q_w q_{w+1} + 78q_w^2\Big] .$$

With the assumption of equidistant points and the homogeneous subsets assumption, the result is derived as in Proposition 5.3.37, but with simplified computation. It is given by

$$h \cdot p_0^{2w}\left[\frac{1}{420}p_0^4 - \frac{47}{840}p_0^3 + \frac{31}{84}p_0^2 + \frac{307}{840}p_0 + \frac{31}{84} - \frac{47}{840}p_0^{-1} + \frac{1}{420}p_0^{-2}\right] .$$

□

Remark 5.3.39. *Although the terms $f_1(p_0, w)$ in Proposition 5.3.37 and $f_1(p_0, \infty)$ in Proposition 5.3.38 are for themselves not easy to interpret, a comparison is much more simple as $f_1(p_0, \infty)$ corresponds to $\lim_{w\to\infty} f_1(p_0, w)$. This is a consequence of an almost identical structure of the evaluated terms. So, the relation of the two terms is easy to interpret by Figure 5.8 where we see that the term in Proposition 5.3.38 corresponds to the lower boundary. For moderate subset level w, however the value of the relation should only slightly differ from one.*

Next, we analyze the general deterministic interpolation error, under the assumption of known derivatives.

Proposition 5.3.40. *The deterministic interpolation error in SuSI with piecewise cubic Hermite spline interpolation with known first derivatives and under equidistant grid points $h = y_{w+1} - y_w$ for all $w = 1, 2, ..., m^* - 1$ is bounded from above by*

$$\int_{y_1}^{y_{m^*}} |s_w^{\text{Spl.}}(y) - q(y)| dy \leq \sum_{w=1}^{m^*-1} h^5 \frac{5}{384} |q^{(4)}(y)|$$

up to the domain approximation error, $\epsilon_{\max}$.

Proof. Follows directly by the error boundary for cubic polynomial interpolation (compare e.g. Hall and Meyer (1976))

$$\max_{y_w \leq y \leq y_{w+1}} |s_w^{\text{Spl.}}(y) - q(y)| \leq \frac{5}{384} \max_{y_w \leq y \leq y_{w+1}} |q^{(4)}(y)| (y_{w+1} - y_w)^4$$

for each of the intervals $[y_w, y_{w+1}]$, $w = 1, 2, ..., m^* - 1$. In total, we thus have

$$\sum_{w=1}^{m^*-1} (y_{w+1} - y_w) \frac{5}{384} |q^{(4)}(y)| (y_{w+1} - y_w)^4 \ ,$$

where $h = y_{w+1} - y_w$ for all $w = 1, 2, ..., m^* - 1$ yields the result. □

Remark 5.3.41. *The deterministic error by interpolation in Proposition 5.3.40 is negligible for smooth functions q. Moreover, the bounds by fourth derivatives are local bounds, so that we receive a tighter boundary by consideration of the domain intervals individually. As a consequence, in particular for lower grid point distances and more sophisticated methods such as 5FD, the bias should become negligible for rather smooth functions. However, we still often favor HM as it does certainly preserve monotonicity of the function q.*

Additionally, we need another lemma which is necessary for the comparison of SuS and SuSI. It links the representation of the variance in SuSI to the failure probability so that a comparison of the variances of SuS and SuSI becomes possible.

Lemma 5.3.42. *With $\mathcal{F}_w^* = \int_{y_w}^{y_{w+1}} f_k^2(y)dy$ the integral of the squared density function with respect to $\mathbf{X}_k$ between the grid points y_w, y_{w+1}. Then, under the homogeneous assumption and for smooth enough conditional failure probability function q or small grid point distances $h = y_{w+1} - y_w$, $w = 1, ..., m^* - 1$, we have*

$$h \sum_{i=1}^{m^*-1} \overline{\mathcal{M}}_{f,p_0} \mathcal{F}_w^* p_0^{2w} f_1(p_0, \infty) + 2h \sum_{1 \le r < t \le m^*-1} \overline{\mathcal{M}}_{f,p_0} \sqrt{\mathcal{F}_r^* \mathcal{F}_t^*} p_0^{r+t} f_1(p_0, \infty) = p_0^{2m}$$

for some

$$\overline{\mathcal{M}}_{f,p_0} \ge \min_{w \in \{1,...,m^*-1\}} \frac{4 \underline{f}_w \underline{s}_w \bar{f}_w \bar{s}_w}{(\bar{f}_w \bar{s}_w + \underline{f}_w \underline{s}_w)^2}$$

and $\underline{f}_w$ the minimum of f_k, $\bar{f}_w$ the maximum of f_k, $\underline{s}_w$ the minimum of $s_w^{\text{Spl.3FD}}$ and $\bar{s}_w$ the maximum of $s_w^{\text{Spl.3FD}}$ on interval $[y_w, y_{w+1}]$.

Proof. Under the homogeneous subset assumption, the real failure probability might be represented in terms of Subset Simulation procedure as

$$p_f = \prod_{i=1}^{m} p_0 = p_0^m \ .$$

On the other hand, in the Subset Simulation Interpolation approach, we have

$$p_f = \sum_{w=1}^{m^*-1} \int_{y_w}^{y_{w+1}} f_k(y) s_w^{\text{Spl.3FD}}(y) dy + \underbrace{\sum_{i=1}^{m^*-1} \int_{y_w}^{y_{w+1}} f_k(y)(q(y) - s_w^{\text{Spl.3FD}}(y))dy}_{=-\text{B}(\hat{P}_f^{\text{Spl.3FD}})}$$

with $\text{B}(\hat{P}_f^{\text{Spl.3FD}})$ the bias by SuSI with 3FD estimation. We then have

$$\sum_{w=1}^{m^*} \int_{y_w}^{y_{w+1}} f_{\mathbf{X}_w}(y) s_w^{\text{Spl.3FD}}(y) dy = p_0^m + \text{B}(\hat{P}_f^{\text{Spl.3FD}})$$

Now, squaring both sides yields

$$\begin{aligned}
&\sum_{w=1}^{m^*}\left(\int_{y_w}^{y_{w+1}} f_{\mathbf{X}_w}(y)s_w^{\text{Spl.3FD}}(y)dy\right)^2 \\
&+2\sum_{1\le r<t\le m^*-1}\left(\int_{y_r}^{y_{r+1}} f_k(y)s_r^{\text{Spl.3FD}}(y)dy\right)\left(\int_{y_t}^{y_{t+1}} f_k(y)s_t^{\text{Spl.3FD}}(y)dy\right) \\
&=(p_0^m+\text{B}(\hat{P}_f^{\text{Spl.3FD}}))^2\ .
\end{aligned} \tag{5.24}$$

Next, by a continuous version of the Pólya-Szegö inequality (Pólya and Szegö (2013), see Dragomir (2003) for a survey on similar type inequalities) we have

$$\begin{aligned}
&\underbrace{\frac{4\underline{f}_w\underline{s}_w\bar{f}_w\bar{s}_w}{(\bar{f}_w\bar{s}_w+\underline{f}_w\underline{s}_w)^2}}_{=:\mathcal{M}_{f,p_0}(w)}\int_{y_w}^{y_{w+1}} f_k^2(y)dy\int_{y_w}^{y_{w+1}}(s_w^{\text{Spl.3FD}}(y))^2dy \\
&\le\left(\int_{y_w}^{y_{w+1}} f_k(y)s_w^{\text{Spl.3FD}}(y)dy\right)^2\ ,
\end{aligned} \tag{5.25}$$

with $\underline{f}_w$ the minimum of f_k, $\bar{f}_w$ the maximum of f_k, $\underline{s}_w$ the minimum of $s_w^{\text{Spl.3FD}}$ and $\bar{s}_w$ the maximum of $s_w^{\text{Spl.3FD}}$ on $[y_w, y_{w+1}]$. Furthermore $\mathcal{M}_{f,p_0}(w)$ is a function depending on f_k and p_0 which can be evaluated for each pair of successive grid points y_w, y_{w+1}. Now, using Equation 5.25 in Equation 5.24 gives

$$\begin{aligned}
&\sum_{w=1}^{m^*-1}\mathcal{M}_{f,p_0}(w)\mathcal{F}_w^*\int_{y_w}^{y_{w+1}}(s_w^{\text{Spl.3FD}}(y))^2dy \\
&\quad+2\sum_{1\le r<t\le m^*-1}\mathcal{M}_{f,p_0}(r)\mathcal{M}_{f,p_0}(t)\sqrt{\mathcal{F}_r^*\mathcal{F}_t^*} \\
&\qquad\cdot\sqrt{\int_{y_r}^{y_{r+1}}(s_r^{\text{Spl.3FD}}(y))^2dy\int_{y_t}^{y_{t+1}}(s_t^{\text{Spl.3FD}}(y))^2dy} \\
&\le(p_0^m+\text{B}(\hat{P}_f^{\text{Spl.3FD}}))^2\ .
\end{aligned}$$

We can next replace $\mathcal{M}_{f,p_0}(w)$ by its minimum $\mathcal{M}_{f,p_0}^{min} := \min_w \mathcal{M}_{f,p_0}(w)$ over all intervals $w = 1, 2, ..., m^* - 1$. Then applying Proposition 5.3.38 and the intermediate value theorem yields the result. □

Before continuing, we would like to explain the result in Lemma 5.3.42 to add a little more meaning, as we would not be able to interpret the result for practical applications later, otherwise.

Remark 5.3.43 (Values for $\mathcal{M}_{f,p_0}(w)$)**.** *First of all, we know that* $s_w^{\text{Spl.3FD}}(y)$ *takes the values* p_0^w *and* p_0^{w+1} *at the grid points for* $w = 1, ..., m^* - 1$*. If the spline does not overshoot those boundaries, those are also maximum and minimum of the spline in this interval, respectively. For the monotone interpolation procedure* $s_w^{\text{Spl.HM}}(y)$ *we never overshoot. 3FD and 5FD interpolation however may overshoot, although this might typically be only minor overshoots. So it is often accurate to replace* $\underline{s}_w = s_w^{\text{Spl.3FD}}(y_w) = p_0^w$ *and* $\bar{s}_w = s_w^{\text{Spl.3FD}}(y_{w+1}) = p_0^{w+1}$*. This allows to rewrite the boundaries of each interval* $w \in \{1, ..., m^* - 1\}$ *by*

$$\frac{4\underline{f}_w \underline{s}_w \bar{f}_w \bar{s}_w}{(\bar{f}_w \bar{s}_w + \underline{f}_w \underline{s}_w)^2} = \frac{4\underline{f}_w p_0^{w+1} \bar{f}_w p_0^w}{(\bar{f}_w p_0^w + \underline{f}_w p_0^{w+1})^2} = \frac{1}{\frac{1}{2} + \frac{1}{4p_0}\frac{\bar{f}_w}{\underline{f}_w} + \frac{p_0}{4}\frac{\underline{f}_w}{\bar{f}_w}} \ .$$

The given boundary however is often not tight at all. Considering local boundaries in the formula instead of the minimum over all intervals and reducing the interval length by choosing higher p_0 *yields a more significant result, but then it still remains unclear which values the correct factor* $\overline{\mathcal{M}}_{f,p_0}$ *in Lemma 5.3.42 takes. In simulations, we however also found many cases where it takes values slightly bigger than one in the dominant regions (where product of pdf and spline value are highest), such as* $1.0 - 1.20$*. In particular for high* p_0 *such results seem to be of a rather stable nature, which is a significant difference with respect to the given sparse boundary. However, this can not be generalized, as it can also be much higher.*

We can now finally compare the variances of SuS and SuSI.

Theorem 5.3.44. *Under the homogeneous subset assumption and for smooth enough* $q(y)$ *or small enough equidistant grid point distances* h *and negligible domain error (compare Lemma 5.3.9), comparison of variance of ordinary Subset Simulation (SuS) and Subset Simulation Interpolation (SuSI) with 3FD yields*

$$\frac{\text{Var}(\hat{P}_f^{\text{Spl.3FD}})}{\text{Var}(\hat{P}_f^{\text{sus}})} \leq \frac{f_1(p_0, \overline{wrt})}{f_1(p_0, \infty)} \overline{\mathcal{M}}_{f,p_0}^{-1} MC^{\text{Spl.3FD}} (1 + p_0^{-m} \text{B}(\hat{P}_f^{\text{Spl.3FD}}))^2$$

for some $\overline{\mathcal{M}}_{f,p_0} \in (\mathcal{M}^{min}_{f,p_0}, \infty)$*, for some* $\overline{wrt} \in (1, m^*-1)$*,* $C^{\text{Spl.3FD}} \in [\frac{1}{m}, \frac{m^*-1}{m}]$*,* $M = \frac{2m^*-m_s}{m}$*, bias given by* $\mathrm{B}(\hat{P}_f^{\text{Spl.3FD}})$ *and* $f_1(p_0, w)$ *as in Proposition 5.3.37.*

Proof. The estimation error with piecewise cubic Hermite interpolation (3FD) is given by

$$\begin{aligned}\hat{P}_f^{\text{Spl.3FD}} - p_f &= \sum_{w=1}^{m^*-1} \int_{y_w}^{y_{w+1}} f_k(y)(\hat{S}_w^{\text{Spl.3FD}}(y) - q(y)) \\ &= \sum_{w=1}^{m^*-1} \int_{y_w}^{y_{w+1}} f_k(y)(\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y)) + \mathrm{B}(\hat{P}_f^{\text{Spl.3FD}})\end{aligned}$$

with bias $\mathrm{B}(\hat{P}_f^{\text{Spl.3FD}})$. First, notice that by Proposition 5.3.37 for every $w = 1, ..., m^* - 1$ the integrals $\mathcal{S}_w = \int_{y_w}^{y_{w+1}} f_k(y)(\hat{S}_w^{\text{Spl.3FD}}(y) - s_w^{\text{Spl.3FD}}(y))dy$ have variance

$$L_w \leq \left(\int_{y_w}^{y_{w+1}} (f_k(y))^2 dy \right) h \mathrm{Var}(\mathcal{E}_w) f_1(p_0, w) \tag{5.26}$$

with $f_1(p_0, w)$ as in Proposition 5.3.37.

Next, as the sum of random variables with zero mean still has zero mean and the variance is given by the sum of the covariances, we have

$$\mathrm{Var}(\hat{P}_f^{\text{Spl.3FD}}) = \sum_{w=1}^{m^*-1} \mathrm{Var}(\mathcal{S}_w) + 2 \sum_{1 \leq r < t \leq m^*-1} \mathrm{Cov}(\mathcal{S}_r, \mathcal{S}_t) \ ,$$

where by Cauchy Schwarz we get

$$\mathrm{Var}(\hat{P}_f^{\text{Spl.3FD}}) \leq \sum_{w=1}^{m^*-1} \mathrm{Var}(\mathcal{S}_w) + 2 \sum_{1 \leq r < t \leq m^*-1} \sqrt{\mathrm{Var}(\mathcal{S}_r)\mathrm{Var}(\mathcal{S}_t)} \ . \tag{5.27}$$

Next, (5.26) allows to bound the variance by

$$\begin{aligned}\mathrm{Var}(\hat{P}_f^{\text{Spl.3FD}}) \leq &\sum_{w=1}^{m^*-1} \mathcal{F}_w^* h \mathrm{Var}(\mathcal{E}_w) f_1(p_0, w) \\ &+ 2 \sum_{1 \leq r < t \leq m^*-1} h \sqrt{\mathcal{F}_r^* \mathcal{F}_t^* \mathrm{Var}(\mathcal{E}_r) f_1(p_0, r) \mathrm{Var}(\mathcal{E}_t) f_1(p_0, t)} \ .\end{aligned}$$

With Proposition 5.3.6, we thus have

$$\begin{aligned}\mathrm{Var}(\hat{P}_f^{\mathrm{Spl.3FD}}) \leq \sum_{w=1}^{m^*-1} h\mathcal{F}_w^* w\sigma^2(\mathrm{SuSI})p_0^{2(w-1)} f_1(p_0,w) \\ + 2\sum_{1\leq r<t\leq m^*-1} h\sqrt{\mathcal{F}_r^*\mathcal{F}_t^*}\sqrt{rt}\sigma^2(\mathrm{SuSI})p_0^{r+t-2}\sqrt{f_1(p_0,r)f_1(p_0,t)}\ .\end{aligned} \tag{5.28}$$

Here, $\sqrt{f_1(p_0,r)f_1(p_0,t)}$ can be replaced by $f_1(p_0,\overline{rt})$ for some $\overline{rt}\in(r,t)$ by the intermediate value theorem.

On the other hand, the variance of SuS, following Proposition 5.3.6, is given by

$$\mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) = m\sigma^2(\mathrm{SuS})p_0^{2(m-1)}\ . \tag{5.29}$$

Now it remains to account for σ^2 in SuS and SuSI as in the staircase approach. Following the approach in Theorem 5.3.18, we again need to include the possibly different number of samples per subset $N^{\mathrm{SuS}}, N^{\mathrm{SuSI}}$ (compare Remark 5.2.3) under fixed total number of total evaluations E_T, having under no-reuse of seeds

$$M^{-1} = \frac{N^{\mathrm{SuS}}}{N^{\mathrm{SuS}}} = \frac{m}{2m^* - m_s}$$

so that $\sigma^2(\mathrm{SuSI}) = M^{-1}\sigma^2(\mathrm{SuS})$. Then (5.28) and (5.29) yield

$$\begin{aligned}\frac{\mathrm{Var}(\hat{P}_f^{\mathrm{Spl.3FD}})}{\mathrm{Var}(\hat{P}_f^{\mathrm{sus}})} \leq p_0^{-2m}\Bigg(\sum_{w=1}^{m^*-1} h\mathcal{F}_w^*\frac{w}{m}p_0^{2w}f_1(p_0,w) \\ + 2\sum_{1\leq r<t\leq m^*-1} h\sqrt{\mathcal{F}_r^*\mathcal{F}_t^*}\frac{\sqrt{rt}}{m}p_0^{r+t}f_1(p_0,\overline{rt})\Bigg)\ .\end{aligned} \tag{5.30}$$

With the aim of converting the term to a similar one as given in Lemma 5.3.42, we rewrite the right hand side in Equation 5.30 as

$$\begin{aligned}p_0^{-2m}\frac{f_1(p_0,\overline{wrt})}{f_1(p_0,\infty)}\frac{1}{\overline{\mathcal{M}}_{f,p_0}}M\Bigg(h\sum_{w=1}^{m^*-1}\overline{\mathcal{M}}_{f,p_0}\mathcal{F}_w^*\underbrace{\frac{w}{m}}_{I}p_0^{2w}f_1(p_0,\infty) \\ + 2h\sum_{1\leq r<t\leq m^*-1}\overline{\mathcal{M}}_{f,p_0}\sqrt{\mathcal{F}_r^*\mathcal{F}_t^*}\underbrace{\frac{\sqrt{rt}}{m}}_{II}p_0^{r+t}f_1(p_0,\infty)\Bigg)\ .\end{aligned} \tag{5.31}$$

for some $\overline{wrt} \in (\min(w,r,t), \max(w,r,t))$.

Now we see that this term is, up to terms I and II, equal to the one in Lemma 5.3.42, resulting in

$$p_0^{-2m} \frac{f_1(p_0, \overline{wrt})}{f_1(p_0, \infty)} \frac{1}{\overline{\mathcal{M}}_{f,p_0}} M C^{\text{Spl.3FD}} (p_0^m + \text{B}(\hat{P}_f^{\text{Spl.3FD}}))^2 \tag{5.32}$$

where $C^{\text{Spl.3FD}} \in [\frac{1}{m}, \frac{m^*-1}{m}]$ accounts for the difference by weighting according to terms I and II.

□

Similar as for the staircase approach, we also add interpretation to the not so easy to interpret result in Theorem 5.3.44.

Remark 5.3.45 (Interpretation of Theorem 5.3.44)**.** *We have*

$$\frac{\text{Var}(\hat{P}_f^{\text{Spl.3FD}})}{\text{Var}(\hat{P}_f^{\text{sus}})} \overset{(*)}{\leq} \underbrace{\underbrace{M}_{(1)} \underbrace{C^{\text{Spl.3FD}}}_{(2)} \underbrace{(1 + p_0^{-m} \text{B}(\hat{P}_f^{\text{Spl.3FD}}))^2}_{(3)} \underbrace{\overline{\mathcal{M}}_{f,p_0}^{-1}}_{(4)} \underbrace{\frac{f_1(p_0, \overline{wrt})}{f_1(p_0, \infty)}}_{(5)}}_{\approx 1?}$$

for some $\overline{\mathcal{M}}_{f,p_0} \in (\mathcal{M}_{f,p_0}^{min}, \infty)$, $C^{\text{Spl.3FD}} \in [\frac{1}{m}, \frac{m^*-1}{m}]$, $M = \frac{m^*-m_s}{m}$ *and* $f_1(p_0, w)$ *as in Proposition 5.3.37. We start with a component-wise interpretation:*

(1) This is the same as in the staircase interpolation approach and accounts for the different number of samples per subset N in SuSI and SuS for same number of total evaluations E_T.

(2) As M_u, M_l, this constant represents the effect by having more weight of the density function in specific grid point intervals.

(3) As in the staircase approach, this corresponds to the different absolute value when a bias is given which induces another coefficient of variation with respect to the fixed true failure probability p_f.

(4) $\overline{\mathcal{M}}_{f,p_0}^{-1}$ is hard to interpret. Remark 5.3.43 offers some support, but sharp boundaries are in general not available. Simulations give evidence that it might be appropriate to assume $\overline{\mathcal{M}}_{f,p_0}^{-1} \approx 1$ in several cases, neglecting the factor, in particular when p_0 is high.

(5) *Figure 5.8 yields a good basis for derivation of possible values of the fraction and strict bounds thereof. Additionally, we can see that in practice it will often approximately be equal to one since* $\overline{wrt}$ *will be rather high. It relates to the importance of subset levels for evaluation of the failure probability so that moderate or high values are typical. Otherwise, the failure probability would be high. In particular for* p_0 *not very low, the difference is low and additionally* m^* *increases so that subset levels which are put more weight on also increase.*

(∗) *In contrast to Theorem 5.3.18, we only have an inequality here. This results from indispensable usage of the Cauchy Schwarz inequality.*

Note that we may represent the bias in the interpolation approach as given in the staircase case by appropriate replacement of c_u *or* c_l *in Theorem 5.3.13, respectively. More details are given in the remainder of this section.*

In comparison to the staircase approach, in this case it is much harder to understand the meaning of the result as we have many unknowns which are also specific with respect to the given particular problem. Furthermore the boundary might not be sparse because of using Cauchy Schwarz type inequalities in the proof (implicitly in (5.26), explicitly in (5.27) and also in a reversed form from (5.31) to (5.32)). The result however still helps to understand the variance in SuSI as we might explore many properties by the proofs and show where the main discrepancies of SuS and SuSI variance are produced and it even is likely that often $\overline{\mathcal{M}}^{-1}_{f,p_0} \frac{f_1(p_0,\overline{wrt})}{f_1(p_0,\infty)} \approx 1$, when p_0 is not too small, so that the formula for splines complies with the one for the staircase approach. In particular, Proposition 5.3.37 can be compared with $\mathcal{F}_w \mathrm{Var}(\mathcal{E}_w) = \left(\int_{y_w}^{y_{w+1}} f_k(y)dy\right)^2 \mathrm{Var}(\mathcal{E}_w)$ which corresponds to the variance of an interval in the staircase approach (compare (5.8)). We just need to evaluate

$$\frac{\left(\int_{y_w}^{y_{w+1}} (f_k(y))^2 dy\right) h f_1(p_0, w)}{\left(\int_{y_w}^{y_{w+1}} f_k(y)dy\right)^2},$$

which was found to be often well approximated by $f_1(p_0, w)$. The information this provides in combination with Figure 5.8 is that the variances are actually similar. Indeed, we see how $f_1(p_0, w)$ typically takes values smaller than one, but the stated comparison relates to the variance of the upper staircase approach. If the lower staircase approach is taken or the average approach,

then we have similar effects that will cancel with $f_1(p_0, w)$. This discovery is in line with our expectations.

The main benefit of the interpolation over the staircase approach comes from the expected bias reduction. Unfortunately, we do not have a formula for the bias which might be practically applied because of the unknown q. Thus, we may either assume q is sufficiently smooth to have an approximately unbiased result or just fix it to specific values for examination of performance. By Proposition 5.3.40 and Proposition 5.3.33 or Proposition 5.3.34, the bias can be bounded from above and becomes small for smooth functions (derivatives of third order $q^{(3)}$ small) or small grid point distances h. If q is smooth enough or h small, the interpolation error can thus become negligible. Still, q is unknown and the bias might as well be relevant. To better account for these different considerations, we utilize another representation of the bias which allows to easily account for the fact that the staircase approach yields boundaries for the bias for any monotone function. Consequently, the shape preserving piecewise cubic Hermite interpolation (corresponding to HM) takes a bias in between the bias of upper staircase and lower staircase approach. To simplify analysis, it is thus admissible to represent the bias by interpolation, similar as in the staircase approach by

$$B = c_I \frac{1 - p_0}{p_0}$$

with $c_I \in [-1, 1]$ under preserving monotonicity with method HM. The rRMSE is then bounded as follows.

Corollary 5.3.46 (rRMSE, Splines)**.** *The* rRMSE *of the estimated failure probability by 3FD is given by*

$$\begin{aligned}\mathrm{rRMSE}(\hat{P}_f^{\mathrm{Spl.3FD}}) \leq \Bigg\{ & MC^{\mathrm{Spl.3FD}} \left(1 + c_I \frac{1 - p_0}{p_0}\right)^2 \overline{\mathcal{M}}_{f,p_0}^{-1} \frac{f_1(p_0, \overline{wrt})}{f_1(p_0, \infty)} \\ & \cdot \frac{(1 - p_0)^r}{p_0 (\log(p_0))^2} \frac{(\log(p_f))^2}{E_T} (1 + \bar{\gamma}) + \left(c_I \frac{1 - p_0}{p_0}\right)^2 \Bigg\}^{\frac{1}{2}}\end{aligned}$$

with $\overline{\mathcal{M}}_{f,p_0}$, $\overline{\mathcal{M}}_{f,p_0}$, $C^{\mathrm{Spl.3FD}}$ *and* $f_1(p_0, w)$ *as in Theorem 5.3.44 and for some* $c_I \in \mathbb{R}$*. The factor* $(1 + \bar{\gamma})$ *accounts for the effective sample number as in Theorem 4.1.11. If seeds are not re-used,* $r = 1$ *and* M *as in Theorem 5.3.44. Otherwise, we have* $r = 2$ *and* M *as in Remark 5.3.19.*

Proof. The proof is derived as in the staircase case. Starting with

$$\mathrm{rRMSE}(\hat{P}_f^{\mathrm{Spl.3FD}}) = \sqrt{\mathrm{Var}(\hat{P}_f^{\mathrm{u}}) + (\mathrm{B}(\hat{P}_f^{\mathrm{Spl.3FD}}))^2}$$

using the derived variance in Theorem 5.3.44 and rewriting the bias we get

$$\frac{1}{p_f} \cdot \left\{ MC^{\mathrm{Spl.3FD}} \left(1 + c_I \frac{1-p_0}{p_0}\right)^2 \overline{\mathcal{M}}_{f,p_0}^{-1} \cdot \frac{f_1(p_0, \overline{wrt})}{f_1(p_0, \infty)} \mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) + \left(c_u \frac{1-p_0}{p_0}\right)^2 \right\}^{\frac{1}{2}}.$$

Using Theorem 4.1.11

$$\mathrm{Var}(\hat{P}_f^{\mathrm{sus}}) = \Big(\mathrm{CV}(\hat{P}_f, E_T)\Big)^2 p_f^2 = \frac{(1-p_0)^r}{p_0(\log(p_0))^2} \frac{(\log(p_f))^2}{E_T} (1+\bar{\gamma}) p_f^2$$

then yields the result. Remark 5.3.19 for consideration of re-using seeds naturally also holds for the spline approach. □

As in Corollary 5.3.46, results for methods 5FD and HM can be derived accordingly. Then the variance should be similar as described in the beginning of this section whereas the bias is bounded for HM by $c_{I,HM} \in [-1, 1]$ and should be lower than in 3FD for smooth functions when choosing 5FD.

5.4 Main Results: Interpretation and Consequences

This section is devoted to a transformation of the theoretical results to practical applications of SuSI and thereby serves as another step towards exploring its real performance. The rRMSE of SuSI was found to depend on many different factors. Although we have already discussed them separately, it still remains hard to understand in which cases SuSI performs good and in which it does not. To improve understanding, we examine the results by statistical analysis for some specific (realistic) scenarios in this section. Because SuS and SuSI might reach their optimal performance under different parameter choices, it is important to not directly compare the two methods for the same p_0. Furthermore, it is important to favor robust parameter choices, so that variations in the outcome of the algorithm should not lead to a high loss in efficiency. To capture all important information, we will compare the rRMSE and intervals containing the realizations of $\hat{P}_f$ with high probability for SuSI and SuS with respect to all $p_0 \in (0,1)$, based on the considered scenarios.

We are not given the shape of the distribution of the failure probability estimators, neither in SuS nor in SuSI. The resulting intervals containing $\hat{P}_f$ with high probability, which we try to approximate, should therefore be viewed critically. Many arguments and conclusions in this section are heuristic ones, since we derived our results under assumption of high sample numbers N, while being most interested in rather low or moderately high N.

Remark 5.4.1. *Even though assumptions such as a Gaussian distribution for the estimation results are violated, the conclusions should remain similar as the assumptions were made for both SuS and SuSI. As long as there is no systematic error involved, favoring one approach over the other by simplifications, the results will not be affected much. Nevertheless, this extrapolation of the results to low N has to be backed by evidence. According to this, simulations are given in Section 6.1.*

For analysis, we set $\overline{\mathcal{M}}_{f,p_0}^{-1} \frac{f_1(p_0,\overline{wrt})}{f_1(p_0,\infty)} \approx 1$ for SuSI with splines, which was found to be admissible in many cases, in particular for high p_0. However, we also apply the formula on low p_0, since this approximation does not necessarily need to suit reality well. There are other good arguments (such as the clear similarity in construction of the estimates and similarities visible in derivation of theoretical results in this section) that variances of staircase

and interpolation approach should be similar and hence not substantially differ. As a consequence, we may then give an approximate upper boundary for the rRMSE of SuSI by the same formula, according to Corollary 5.3.22 and Corollary 5.3.46 in the staircase and spline approach, respectively.

Remark 5.4.2 (Formula)**.** *We use the following formula for evaluation (compare Corollary 5.3.22 and Corollary 5.3.46)*

$$\begin{aligned}\mathrm{rRMSE}(\hat{P}_f^*) \leq \Bigg\{ M(p_0) C \left(1 + c\frac{1-p_0}{p_0}\right)^2 \frac{(1-p_0)^r}{p_0(\log(p_0))^2} \frac{(\log(p_f))^2}{E_T}(1+\bar{\gamma}) \\ + \left(c\frac{1-p_0}{p_0}\right)^2 \Bigg\}^{\frac{1}{2}},\end{aligned} \tag{5.33}$$

where we replace M by $M(p_0)$ to underline its dependency on p_0 when subset seeds are re-used. $\hat{P}_f^*$ refers to the estimator by any interpolation method. C and c are written without superscript or subscript in contrast to the previous analysis, respectively, since we jointly analyze interpolation and staircase approximation.

The corresponding c in Remark 5.4.2 is unknown but bounded by $c \in [-1, 1]$ for monotone interpolation functions. For smooth conditional failure probability functions q, c becomes small $c \approx 0$. Note that the formula only yields an upper bound for the rRMSE. In reality it might as well perform better, in particular because the boundary was derived under usage of the Cauchy Schwarz inequality.

Next, we define scenarios for our analysis. The scenarios, which require to define C and M, are created as follows. For factor M which describes the effect by different N under the same number of total evaluations E_T for SuS and SuSI, we use SuSI with a maximum error of approximately $2\% \cdot p_f$ so that we can assume two more subsets on average for $p_0 = 0.1$ (1% error on each side of the domain of $\mathbf{X}_k$) and $2\frac{\log(0.1)}{\log(p_0^*)}$ more subsets for any $p_0^* \geq 0.1$.

Remark 5.4.3. *Instead of the number of total evaluations, it could also be appropriate to count the number of sample creations N_T instead, if either we indeed want to reduce the sample number, have meta models that allow for cheap evaluation or know the ordering of the samples according to the changing limit state function with respect to $\mathbf{X}_k$, e.g. when it is kept in successive subsets. In this case, SuSI benefits a lot, practically requiring only half of the*

samples (compare Remark 5.2.3), which is equal to ordinary SuS. Then M is significantly decreased.

The maximum error can often be chosen higher, since we do not need to be very accurate in the deterministic error, when the variance is much higher under small sample numbers. Also note that the relative increase of subsets remains the same for any p_0^* under same p_0 for initial SuS and successive SuSI steps, so that the effect does not change significantly for different M, except for higher variation when selecting lower p_0. This simplification is admissible, since we can adapt our N's in each subset to suit the corresponding p_0. Another point that defines M is the number of subsets which are created by SuS steps before the initial SuSI step. Those do not need the double evaluations which are typically required in SuSI steps. By Remark 5.2.3, or as explicitly restated in Theorem 5.3.18 and Theorem 5.3.44 we can derive results when not re-using seeds or as in Remark 5.3.19 under re-use of seeds.

Example 5.4.4. *If we have a failure probability of $p_f = 5 \cdot 10^{-7}$, then we need 7 subsets in SuS and $9 = 7 + 2$ subsets in SuSI with maximum error 2% for $p_0 = 0.1$ on average. Then, under including the effect of SuS steps before SuSI steps, depending on the significance of $\mathbf{X}_k$ for the failure probability, we could have*

$$M = \frac{2m^* - m_s}{m} = \frac{2 \cdot 9 - 6}{7} \approx 1.7$$

for mostly irrelevant $\mathbf{X}_k$ and

$$M = \frac{2m^* - m_s}{m} = \frac{2 \cdot 9 - 2}{7} \approx 2.3$$

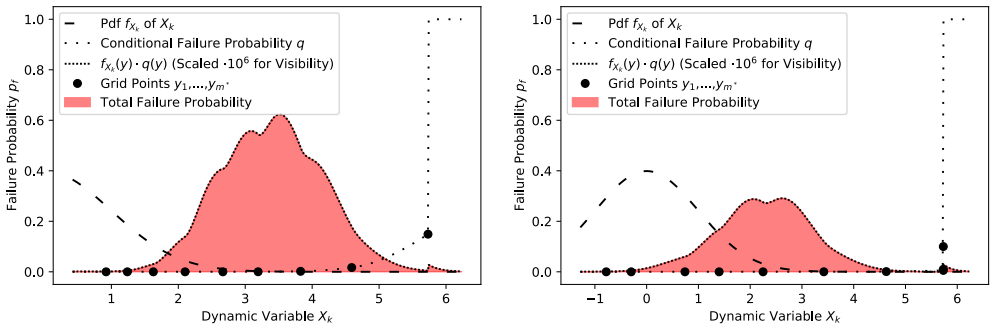
when $\mathbf{X}_k$ is the dominant variable. Under re-use of seeds, this changes to

$$M = \frac{m^*(2 - p_0) - m_s}{m(1 - p_0)}$$

dependent on p_0. Concerning C, we assume $C \approx 0.5$ for the dominant variable case and $C \approx 1$ when $\mathbf{X}_k$ is not so relevant. An illustration for justification of the plausibility of this assumption is given in Figure 5.9.

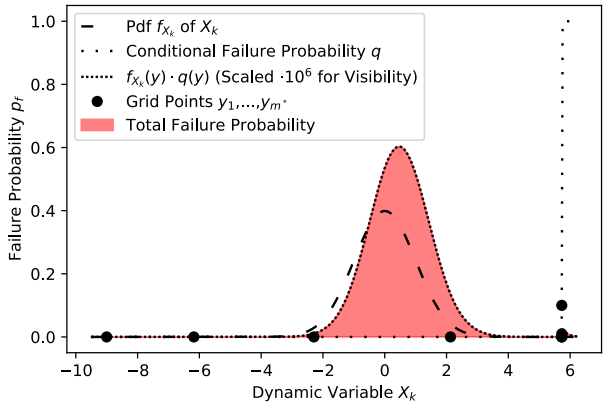
Considered scenarios, resulting from relevance of $\mathbf{X}_k$ and varied number of total evaluations, are stated in Table 5.1. In total, we then have 9 scenarios, trying to cover enough cases to get an intuition of the effectiveness of SuSI

Figure 5.9: Illustration of the effect of different significances of the dynamic variable $\mathbf{X}_k$ on the placement of the most important parts in derivation of the total estimated failure probability $\hat{p}_f$ on the domain of $\mathbf{X}_k$. The difference is best recognized by the relative placement of the important evaluation parts of the domain with respect to the pdf of $\mathbf{X}_k$. The procedure follows Figure 5.1, where again $f_{\mathbf{x}_k}(y) \cdot q(y)$ was scaled so that it becomes clearly visible, but instead the standard limit state function g_1 (compare Equation 4.2) was considered.



(a) Dominant $\mathbf{X}_k$, $d = 2$

(b) Moderate $\mathbf{X}_k$, $d = 5$



(c) Insignificant $\mathbf{X}_k$, $d = 100$

Scenario	Description		Model Parameters		
Short name	Role of $\mathbf{X}_k$	$\#N$	C	M	E_T
nD-Nl	Insignificant	low	1.0	1.7	$3 \cdot 10^3$
nD-Nm		moderate			$1 \cdot 10^4$
nD-Nh		high			$8 \cdot 10^4$
M-Nl	Moderate	low	0.75	2.0	$3 \cdot 10^3$
M-Nm		moderate			$1 \cdot 10^4$
M-Nh		high			$8 \cdot 10^4$
D-Nl	Dominant	low	0.5	2.3	$3 \cdot 10^3$
D-Nm		moderate			$1 \cdot 10^4$
D-Nh		high			$8 \cdot 10^4$

Table 5.1: Considered scenarios for analysis, based on a calibration with a real example. nD, M and D refer to the importance of $\mathbf{X}_k$ and Nl, Nm, Nh to the number of samples, in ascending order. The failure probability was set to $p_f = 5 \cdot 10^{-7}$ and the dependency factor was set to $\bar{\gamma} = 3.0$ in all cases. Note, M^* depends on p_0, as given in Example 5.4.4, if seeds are re-used, so that the corresponding table values do not necessarily represent such cases, relying on a constant M^*.

in different realistic settings. For plotting the rRMSE, we can just plug the corresponding scenario parameters into Equation 5.33, whereas for the intervals containing $\hat{P}_f$ with high probability, we evaluate bias and variance in Equation 5.33 individually and then state the intervals of the corresponding quantiles of the distribution of the estimates. To do so, we have to assume a distribution of the estimator $\hat{P}_f$, where we choose the Gaussian or lognormal distribution. Appropriate choices depend on N, where for small N these are more likely to take a lognormal shape because they are generally heavy-tailed. For high N, we apply a Gaussian distribution. To define the intervals containing $\hat{P}_f$ with high probability, we use statistically derived bias and variance to define the first two moments and thereby, under assumption of a distribution type, derive the distribution of the result. Note that it is not a priori clear, if the resulting rRMSE, and thus bias and variance, can be extrapolated to low N as it is derived under asymptotic arguments.

Before showing the results, we want to remind of our focus on the dynamic model application.

Remark 5.4.5 (Interpretation: Use-Case, n Constellations)**.** *We often set*

the number of relevant constellation equal to one ($n = 1$) for comparison as stated in Section 5.1.1, whereas SuSI is not designed to perform better at such static models. Instead SuSI focuses on dynamic models with $n > 1$. So if SuSI is worse for $n = 1$, we know that SuSI takes a higher computational demand than a corresponding static reliability evaluation by SuS but of course would still typically outperform repetitive SuS evaluations in the dynamic model setting for higher n. The number of constellations n is thus also varied to really capture the benefits by SuSI. Then it is assumed that the number of total evaluations E_T is equally distributed across the individual SuS runs so that SuS is considered to have only E_T/n total evaluations for evaluation of every constellation, in contrast to E_T in SuSI where one run delivers the result for all n constellations. This is a questionable way of comparing efficiencies, since the use-cases of SuS are static reliability models and extensions for more dynamic model variants, as discussed in Section 3.1, exist. However, it is not clear which method to use for comparison and often methods for dynamic models are based on intractable approximations (e.g. when utilizing surrogate models or local Taylor approximations) which SuSI is not. Furthermore, ordinary SuS is very popular with known computational demands so that information relating to SuS offers an easily understandable interpretation. The proposed procedure can also be interpreted as comparing the computational demands of SuSI with n static reliability evaluations by SuS. Finding a number of constellations n where SuS then provides the same rRMSE *as SuSI corresponds to SuSI requiring computational demands as low as n static reliability evaluations by SuS for evaluation of a dynamic model.*

In the subsequent studies, it is important to be aware of the number of model constellations n, as stated in Remark 5.4.5, to judge the given results. Also note, that utilization of SuS for interpolation instead, is not straightforward nor efficient. Instead of computing all relevant constellations by a SuS result, one could also think of spanning the domain of the dynamic variable $\mathbf{X}_k$ with grid points. To do so, one would require a SuS evaluation for each grid point. This is impractical because of three main reasons.

- This approach can be infeasible for complex structures as many grid points are required, in particular if the maximum bias should be small.
- It is not clear, how to find appropriate grid points at all. The prediction step in SuSI is crucial to do so and we would need something similar for SuS, also including the potential drawbacks of SuSI this way.

- The result will generally have a higher variation. Every SuS estimation is individually exposed to stochastic noise. We might thereby receive wrong results, even with respect to monotonicity of q, in general.

Also, it is important to remember that a twofold rRMSE does not induce double computational demands.

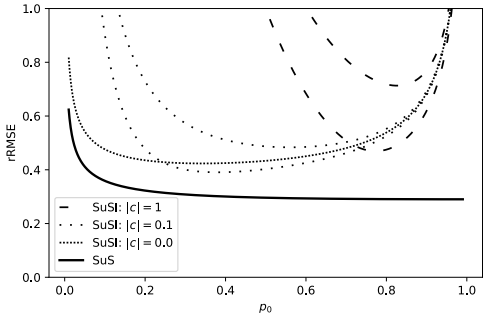
Remark 5.4.6. *Convergence of SuS and SuSI is $O(1/N)$. As a consequence, we need to approximately quadruple the number of total evaluations E_T, to achieve half of the standard deviation, which is contained in the* rRMSE *and shown in the scenario illustrations.*

The capability of SuSI to compute failure probabilities with respect to several constellations at once, additionally provides the following benefit.

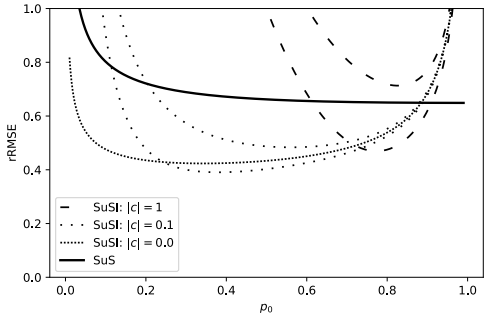
Remark 5.4.7 (Relations of Failure Probabilities)**.** *When using SuSI for derivation of several model constellations, e.g. deriving time-dependent constellations by one run, relations of the corresponding failure probabilities are calculated particularly efficient. In comparison to ordinary SuS, we then have a reduction of stochastic noise in these relations because all results are derived by the same simulation run.*

Figure 5.10 shows the rRMSE for several constellations, demonstrating that SuS is often more efficient than SuSI for $n = 1$ but at the same time showing that SuSI typically outperforms repetitive SuS, or equivalently does not require more computational demands than a few static SuS evaluations, in the use-case, as demonstrated by setting $n = 5$. In addition we see that it is important to keep track of the bias in SuSI as it might decrease its performance significantly, even taking worse results than SuS in case of $n = 5$. In addition to the statistical derivations of the bias, having $c = 1$ for upper and $c = -1$ for lower staircase approximations, we also examine the cases $c = 0.1$ and $c = -0.1$ which is not a boundary anymore but could be considered as more realistic values, expecting the conditional failure probability function to not fluctuate extremely. Also the case of unbiased SuSI estimation $c = 0$ is considered for direct comparison of the variances. In general, SuSI performs better for more relevant $\mathbf{X}_k$. This effect however is exceeded by the number of relevant constellations n. For small n, we see that a highly positive bias as for the upper staircase case results in the worst performance. However, in contrast to this bad performance, it yields security with respect to systematic underestimations of the result.

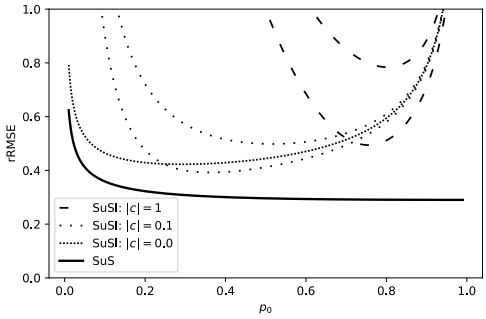
Figure 5.10: Comparing the efficiency of SuS and SuSI for different scenarios. A moderate sample number is assumed, considering a single reliability evaluation $n = 1$ and $n = 5$ evaluations. The subset seeds are assumed to be re-used for evaluation (implementation (I4*)).



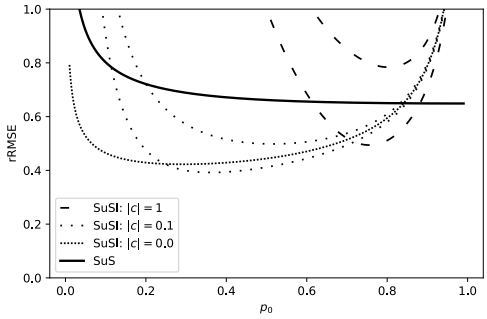
(a) Scenario nD-Nm, $n = 1$



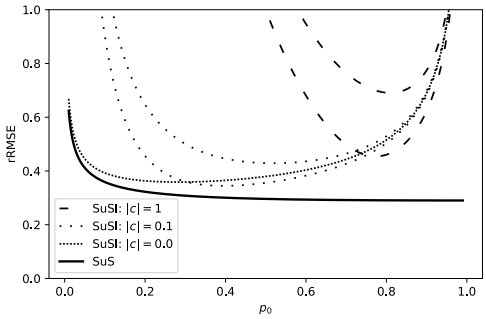
(b) Scenario nD-Nm, $n = 5$



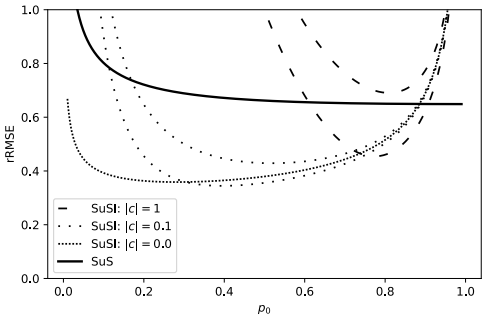
(c) Scenario M-Nm, $n = 1$



(d) Scenario M-Nm, $n = 5$



(e) Scenario D-Nm, $n = 1$

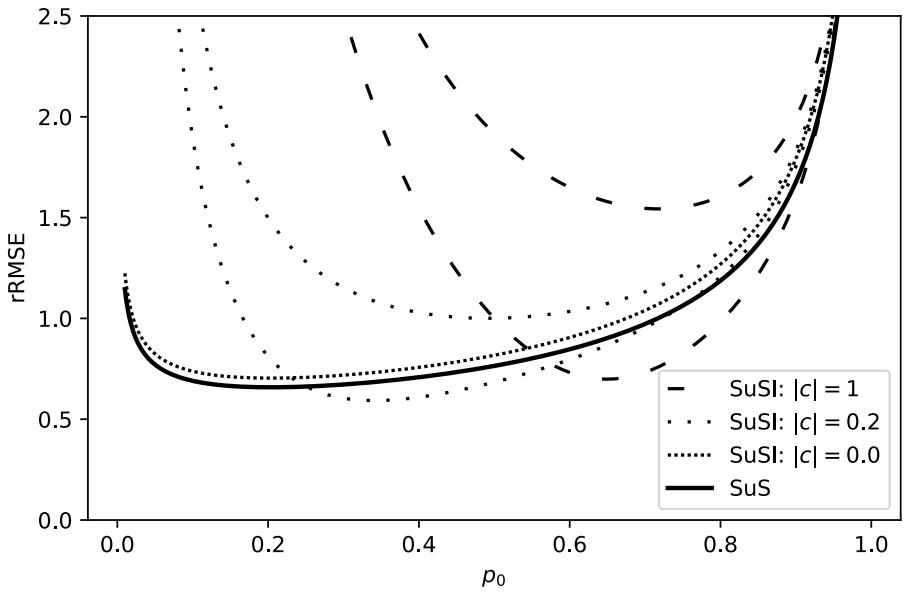


(f) Scenario D-Nm, $n = 5$

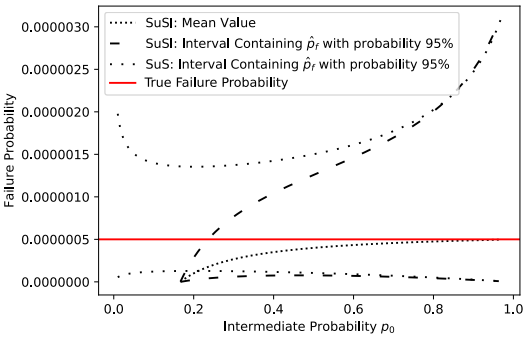
Remark 5.4.8 (Optimal Choice of p_0)**.** *The optimal choice of p_0 is not clear, but moderate values appear to achieve a good balance between total sample number and bias. In general, as the bias is unknown, one has to make the decision if the optimal p_0 should be chosen so that the maximal bias of the result is minimized, or minimize the* rRMSE *with respect to the most plausible bias. Also note that for higher p_0, the bias becomes smaller which allows for more unlikely and smaller systematic errors. In all cases, the real bias can be bounded but remains unknown, demanding a careful interpretation of the results. In contrast to the interpolation approaches, we should certainly reach $c \approx 0$ for approaches by regression with many repetitive runs of the algorithm, even if p_0 is chosen small.*

In Figure 5.11, we demonstrate how SuSI can even have a higher efficiency for a single evaluation according to statistical analysis ($c = -0.2$ negative bias). This is surprising because we did derive an upper boundary for the variance of SuSI and utilized approximations that favored SuS so that real simulations should give even better results for SuSI. Therefore, it possibly outperforms SuS noticeably even in the single evaluation case which is not even considered a use-case for SuSI. The negative bias reduces the absolute mean value of the estimation under approximately constant relative variation. Although this tends to underestimate the result, it might still be useful to take the approach because of higher confidence on results. However, we also have to be careful with results corresponding to a negative bias and high rRMSE. An rRMSE = 1 can always be achieved by setting $\hat{P}_f = 0.0$ constant. Then the relative bias is certainly one, regardless of the real failure probability p_f, and the variance is zero. Thus rRMSE = 1 is a critical benchmark and provides the reason for the behavior of the interval probabilities containing $\hat{p}_f$ with probability 95% of SuSI in Figure 5.11b and Figure 5.11c. Still, the situation of a slightly negative c performing best, is a welcome discovery. Indeed we know that shape-preserving piecewise cubic Hermite interpolation (approach HM) tends to slightly underestimate the result (see Remark 5.3.24). Although we have to be cautious, as an underestimation can become dangerous, it is interesting that this property might induce a decreased rRMSE for our favorite approach. Extensive simulation studies are required, where we also need to examine how the outcome is distributed in reality. Then we may judge better, if it yields a potential benefit instead of a drawback. In contrast to the best performance for low N, the bias becomes less acceptable for high N and efficiency is reduced for insignificant $\mathbf{X}_k$ as

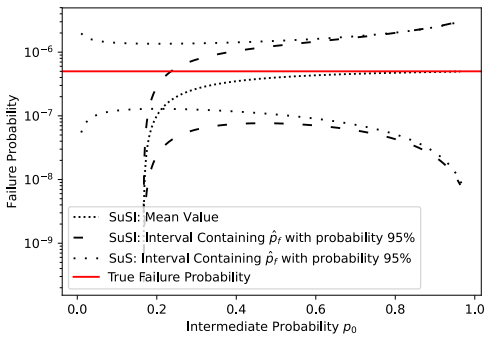
Figure 5.11: Comparing the efficiency of SuS and SuSI for Scenario D-Nl under no re-use of seeds and for a single reliability evaluation $n = 1$. We consider the rRMSE as well as the corresponding intervals containing realizations of $\hat{P}_f$ with probability 95% (quantiles 2.5%, 97.5%), under assumption of lognormally distributed $\hat{P}_f$. In Figure 5.11b and Figure 5.11c, we let $c = -0.2$.



(a) Scenario D-Nl



(b) Scenario D-Nl, linear scale, $c = -0.2$



(c) Scenario D-Nl, log scale, $c = -0.2$

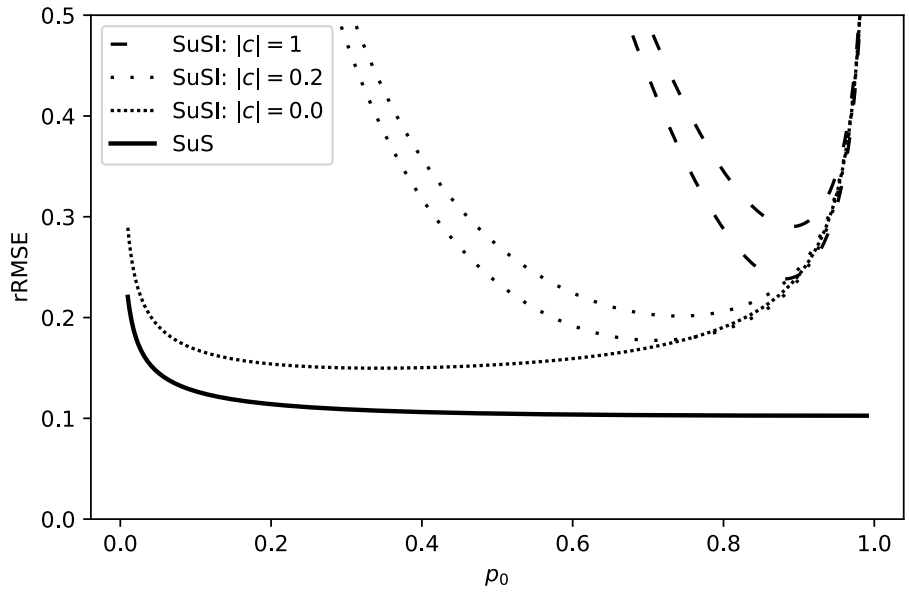
shown in Figure 5.12.

Next, we additionally consider higher number of relevant constellations $n = 10$ and $n = 50$ in Figure 5.13. We look at scenario M-Nh, having a moderate impact of $\mathbf{X}_k$ on the failure probability. In this case, it is clear that we benefit a lot by using SuSI. Although it might also become interesting to use different methods instead of repetitive SuS here. This comparison aims at examples where SuS performs best and different constellations of the model can not reasonably be derived by local approximations or do not provide good information on approximation errors. An example, where potentially many relevant constellations are present, is when the distribution of $\mathbf{X}_k$ is unclear and several cases need to be evaluated. For $n = 50$, the rRMSE of SuS is approximately $3 - 5$ times lower than that of SuSI in most cases. To achieve the same accuracy in SuS thus requires $9 - 25$ times as many computational demanding limit state evaluations as it does when using SuSI. Thus SuSI performs orders of magnitude better in cases, where reliability has to be evaluated with respect to many different constellations of the dynamic variable.

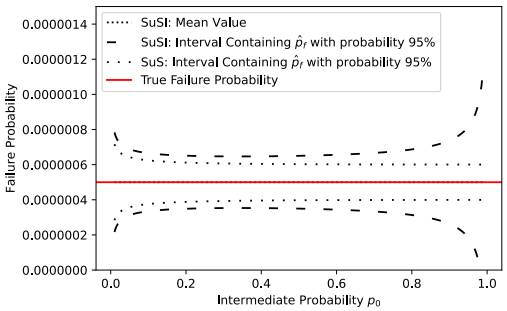
Remark 5.4.9. *The best benefit of SuSI is achieved for high numbers of constellations n and dominant dynamic variable $\mathbf{X}_k$, where it is worthwhile that the corresponding number of evaluations necessary for SuS to receive the same accuracy is often infeasible. In particular, when one has to test several distribution functions to give a clear picture of the unknown effects of the dynamic variable on the failure probability, it is also questionable whether interesting cases will even be explored or if many details just remain unseen because of avoiding the extraordinary high computational costs and basing results on a few single evaluations. Thus, defining n does not capture the relevance of the SuSI approach fully. In contrast, there are cases where such an extensive analysis by SuSI might be irreplaceable for the desired information. In Chapter 8 we will also see how the gain in information can help to create a more complete interpretation of reliability analysis.*

Also remarkable is the opportunity to prevent the drawback of a given bias by either using additional evaluations conditional on important regions of $\mathbf{X}_k$ after termination of the algorithm or by regression such as fitting a smoothing spline to the points of several repetitive independent runs of SuSI. This allows for approximately unbiased results, mostly independent of q, and requires only slightly more limit state evaluations than the analyzed interpolation methods. Anyhow, smoothing splines or regression approaches

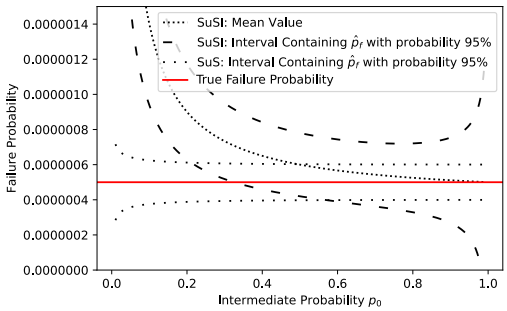
Figure 5.12: Comparison of the efficiency of SuS and SuSI for Scenario nD-Nh under no re-use of seeds for a single reliability evaluation $n = 1$. We consider the rRMSE as well as the corresponding intervals containing realizations of $\hat{P}_f$ with probability 95%, under assumption of lognormally distributed $\hat{P}_f$. In Figure 5.12b and Figure 5.12b, we let $c = -0.2$.



(a) Scenario nD-Nh

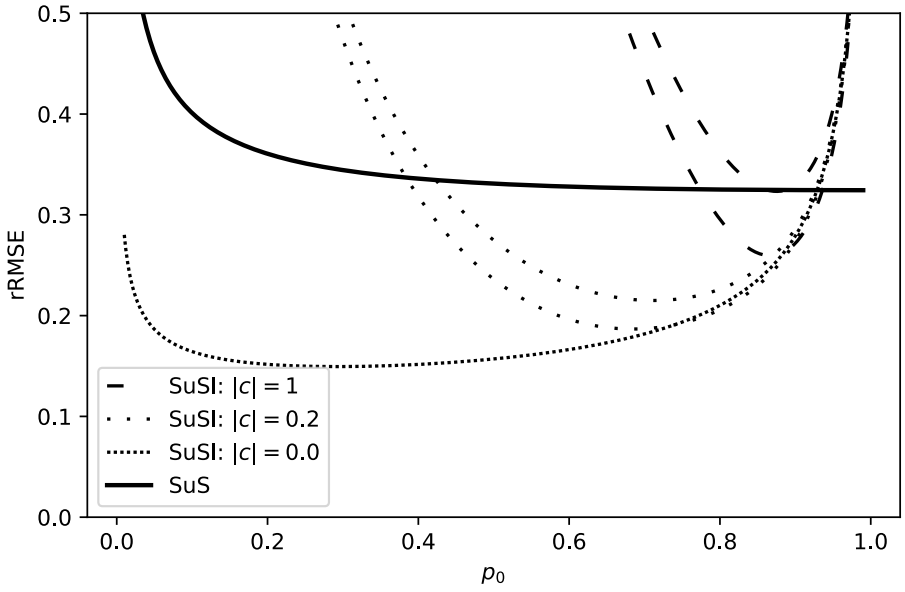


(b) Scenario nD-Nh, $c = 0.0$

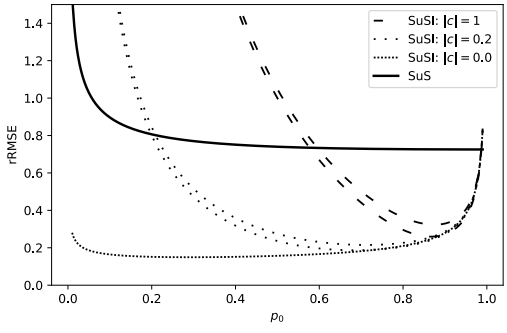


(c) Scenario nD-Nh, $c = 0.2$

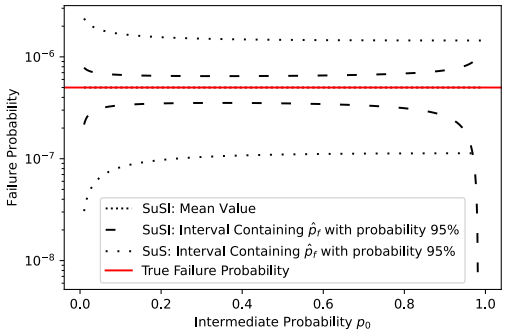
Figure 5.13: Comparison of the efficiency of SuS and SuSI for Scenario M-Nh under re-use of seeds for $n = 10$ and $n = 50$ reliability evaluations. For the intervals containing realizations of $\hat{P}_f$ with probability 95%, we assume SuSI normally distributed due to the high sample number and the single SuS evaluations to be lognormally distributed due to their small corresponding sample number.



(a) Scenario M-Nh, $n = 10$



(b) Scenario M-Nh, $n = 50$



(c) Scenario M-Nh, $n = 50$, $c = 0.0$, log scale

in general are not analyzed here because that would go beyond the scope of this work, needing to calibrate many parameters and requiring repetitive runs of SuSI for a single evaluation.

Chapter 6

Subset Simulation Interpolation: Simulation Study and Applications[1]

In this chapter, after having extensively analyzed SuSI theoretically, we now examine it practically by simulation, thereby giving evidence for our claims of the previous chapter by testing the true performance of the algorithm directly. Furthermore, some application opportunities of SuSI, such as time-dependent reliability evaluation, increased robustness (ergodicity examination, see Remark 3.2.2) of SuS results and sensitivity analysis are presented.

6.1 Simulation Study

We use modern MCMC techniques with constant MCMC chain length of $N_l = 10$ (implementation (I4*), see Chapter 4, Table 4.1) to compare the efficiency of SuSI and SuS. This is in favor of SuS, where using traditional implementations (I1,I2) would favor SuSI according to Remark 5.3.21 (also compare Blandfort et al. (2019a)). Our examples will relate to cases where adaptive sampling works particularly well, which also is in favor of ordinary SuS with respect to comparison of their efficiency, since the within subset sample dependencies will not increase much with respect to the subset level.

[1]The algorithm was first presented at ESREL2019 (Blandfort et al. (2019a)). Since then it has been further developed, applications were extended and a theoretical foundation was given. The result is presented in this dissertation.

Remark 6.1.1 ($n = 1$)**.** *For ease of understanding, we fix the number of constellations to $n = 1$ always. Again note that this only aims at demonstrating the general efficiency of SuSI being similar to that of SuS, allowing to evaluate dynamic models instead of the considered static ones at this cost. So, even if for the single constellation case SuSI performs significantly worse, we will often still favor it over SuS for dynamic models.*

As described in Chapter 2, we will use the rRMSE for analysis of efficiency. It should be pointed out that the rRMSE weights higher distances from the real p_f much stronger than small ones, following a quadratic increase with respect to the distance. This is not generally an issue, but for low N we are having a heavy tailed distribution of the estimators in SuS and SuSI where we might not want to penalize rare overestimations of the failure probability extremely, which we do by applying the rRMSE. Also comparing performances for different number of total samples E_T is thus not straightforward and an increased sample number will most likely increase the efficiency more than expected because of the beneficial change in the shape of the distribution of $\hat{P}_f$, i.e. it will become less heavy-tailed. As a consequence, and as the distributions of SuS and SuSI might substantially differ for small N, we consider supplementary information.

Remark 6.1.2 (MAE, Quantiles)**.** *In addition to the* rRMSE*, we also consider the mean absolute error (MAE) and quantile values. As we propose a new algorithm, it is crucial to give a more full examination of the shape of the result. So, the combined information of* rRMSE*, MAE and quantiles will also help us to understand the algorithm better and allows us to not rely on a single loss function completely, although we still favor the* rRMSE *for measuring the performance and pay the most attention to it.*

With regards to the bias, we want to draw attention to an interesting point. In the presented simulations, the grid points are not fixed, as they are also stochastically drawn in a regular simulation in practice. Thus the bias might differ from the bias that would be given just as the bias by interpolation concerning the smoothness of q derived in the statistical analysis part. Remark 6.1.3 shortly states the consequence of such a grid point selection.

Remark 6.1.3 (Stochastic Grid Point Selection)**.** *By the adaptive creation of the grid points, the bias of repetitive SuSI runs is a combination of several bias, depending on the individual grid point selections. In general, this should*

reduce the probability of worst case scenarios for the bias. Parts of the underlying conditional failure probability function q that behave erratic, will be spanned by the grid points differently in independent simulation runs. This allows to avoid worst case scenarios in many cases, so that the bias might typically not take exceptionally high values when averaging over repetitive runs.

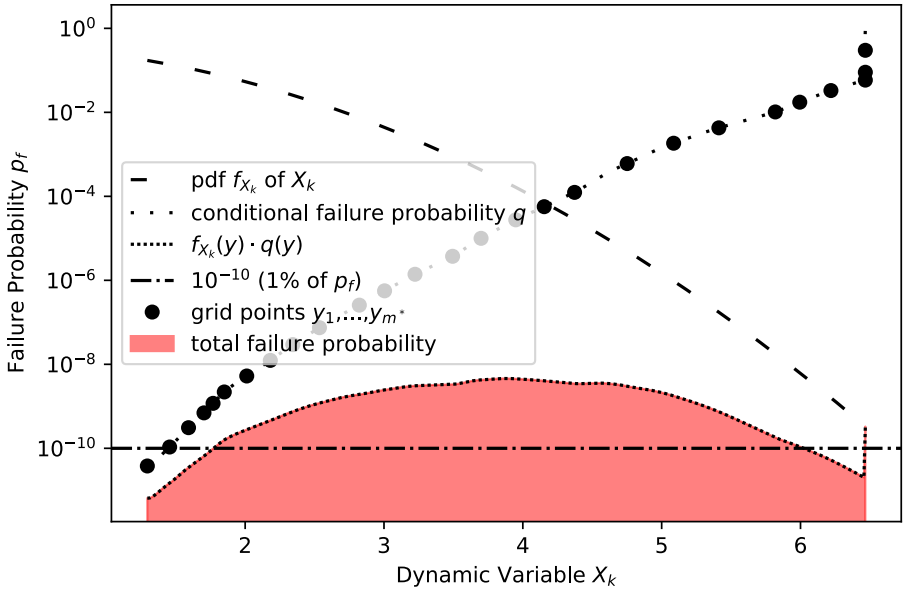
Before we give results by simulation, we want to draw attention to the underlying difference of the studied limit state functions. We again consider g_1, g_{2a} and g_{2b} (compare Equation 4.2-4.4) as a basis for our analysis since these differ to a great extent in their characteristics and cover extreme cases as recognizable in Figure 6.1 and Figure 6.2 . In particular, the big difference in the shape of the conditional failure probability function q is important. Consequently, studying these limit state functions allows to draw conclusions on many settings. At the same time, defining an appropriate adaptive selection of the grid points in the prediction step to well suit all these settings at the same time is a challenging task. It might even be used to test adaptively defined prediction step implementations for provision of general applicability. In our examples, we found that after successfully testing our implementation with these limit state functions, SuSI also worked well out of the box for other more realistic examples, such as in Section 6.2. Figure 6.1c now provides a good basis for understanding and reasoning on results obtained by simulation.

Remark 6.1.4 (Bias by SuSI Estimation)**.** *According to Figure 6.1 and Figure 6.2, a high bias would require the conditional failure probability function to frequently behave erratic at particular points. Even if the function is not smooth, only very specific functions will result in a high relative bias.*

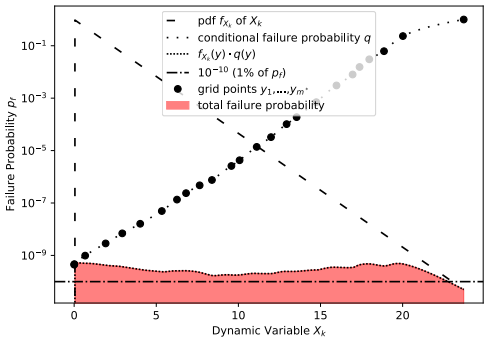
For insightful simulation studies, we need to artificially alter limit state functions for consideration of cases with important and with less important dynamic variables.

Remark 6.1.5 (Construction of Different Limit State Functions for Testing)**.** *To consider different cases of important or non import dynamic variables, we vary the dimension of the limit state functions g_1, g_{2a} and g_{2b}. The used implementation (*I4*) *is rather insensitive with respect to the shape of the failure domain, so that although there might be an impact on the failure domain, simulation results should at least remain similar. This is suspected*

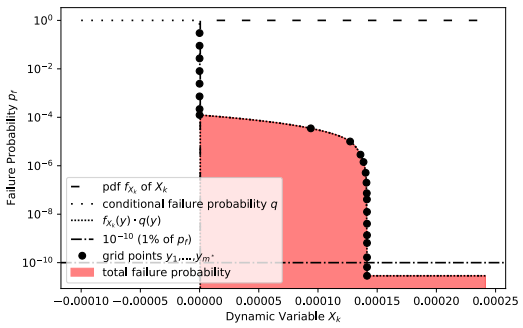
Figure 6.1: Illustration of the effect of considering different limit state functions g_1, g_{2a} and g_{2b} on the relevance of the domain of $\mathbf{X}_k$ with respect to the total estimated failure probability p_f. Typical results by application of SuSI are shown. The dimension is chosen to be $d = 2$ (high impact of $\mathbf{X}_k$).



(a) g_1, $\mathbf{X}_k \sim \mathcal{N}(0,1)$

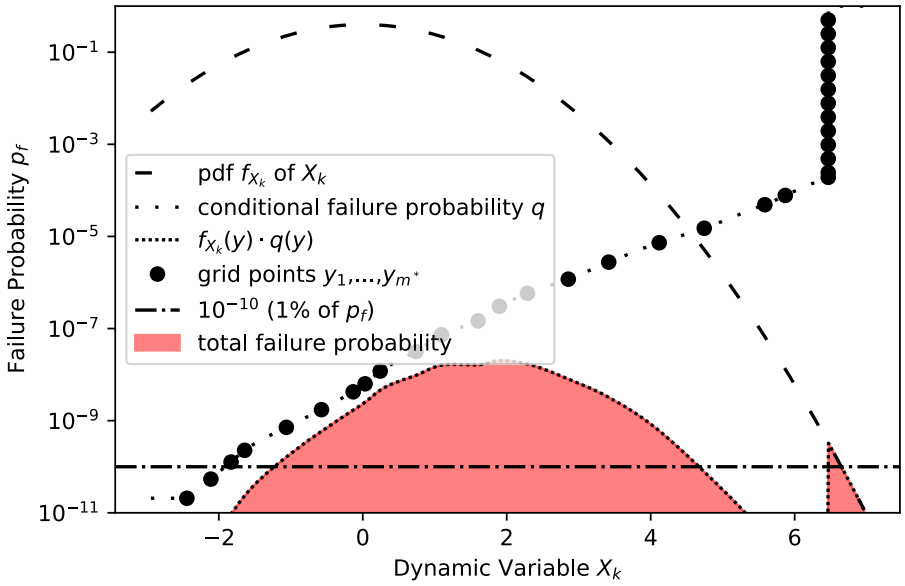


(b) g_{2a}, $\mathbf{X}_k$ exponentially distributed with parameter 1.

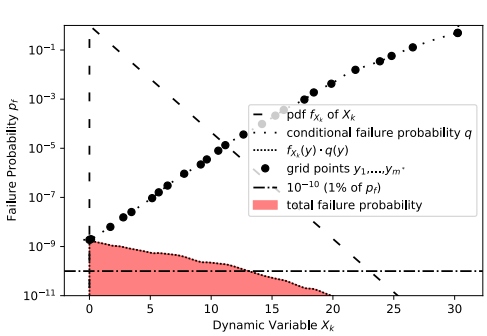


(c) g_{2b}, $\mathbf{X}_k$ exponentially distributed with parameter 1.

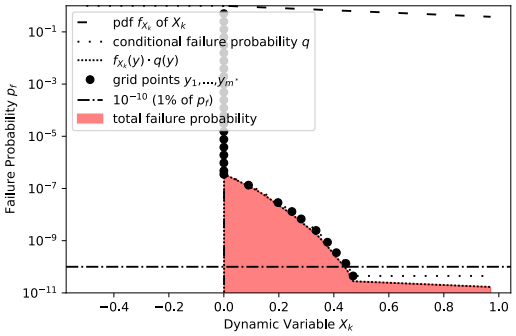
Figure 6.2: Illustration of the effect of considering different limit state functions g_1, g_{2a} and g_{2b} on the relevance of the domain of $\mathbf{X}_k$ with respect to the total estimated failure probability p_f. Typical results by application of SuSI are shown. Dimension is chosen to be $d = 10$ (low impact of $\mathbf{X}_k$).



(a) g_1, $\mathbf{X}_k \sim \mathcal{N}(0,1)$



(b) g_{2a}, $\mathbf{X}_k$ exponentially distributed with parameter 1.



(c) g_{2b}, $\mathbf{X}_k$ exponentially distributed with parameter 1.

in Papaioannou et al. (2015) and also supported by our own studies. The main object of interest, is the effect of the conditional evaluation so that this altering of the limit state functions suffices our demands. We consider dimension $d = 2$ for important $\mathbf{X}_k$ and $d = 10$ for the case of less relevant $\mathbf{X}_k$. Because the single variables of the considered limit states are equally affecting the failure probability, $\mathbf{X}_k$ is set to an arbitrary one of those, without loss of generality. Such low dimensions might not be considered a use-case for application of Subset Simulation in general, but nevertheless results are expected to be similar for high dimensions so that we get an informative and feasible simulation study this way.

Remark 6.1.6 (Simulation Settings)**.** *In the presented simulation studies, the chain length in MCMC is set to $N_c = 10$ in all cases (implementation (I4*)). This allows for a coefficient of variation $\mathrm{CV}(\hat{P}_f)$ being approximately independent of p_0 (compare Section 4.1.3), to increase the efficiency for higher choices of p_0, which may decrease the bias in SuSI. Beyond the choice of $[p_l, p_u]$ for the SuSI-like steps, we also require to set an intermediate probability p_0 for initialization by SuS. The choice of this p_0 is not important for efficiency, but it should not be too low so that one avoids big grid point steps that result in missing evaluation of some important parts of the domain of $\mathbf{X}_k$ more densely. We fix $p_0 = 0.5$ without much loss of generality. Also, when using version a) of the prediction step, we minimize the samples for prediction of appropriate grid points by choosing a rather large interval length, such as $|p_u - p_l| = 0.2$. So, after the prediction step, we accept a candidate grid point with expected intermediate probability p for some $p \in (p_l, p_u)$ which can be substantially different with respect to the subset level. To still preserve good efficiency, the corresponding N have to be adapted on the fly. We set $N = N^* \cdot (\log(p)/\log(p^*))$ for some fixed p^*. This way, the number of samples is distributed efficiently across subsets with potentially different intermediate probability estimates. In version a) of the prediction step, we also reduce the number of samples in the prediction step when higher p_0 are used to keep the total number of samples similar. However, we do not reduce it enough to get an equal number of total samples as it should not become too small, needing to use more samples for prediction if higher intermediate probabilities are chosen. For interpolation (Algorithm 2 A2-5), which is also utilized for step b) (Algorithm 2 A2-2), we use shape preserving piecewise cubic Hermite interpolation (compare Section 5.3). Function values beyond the evaluated domain are assumed to take the highest failure probability pos-*

sible, using the worst case staircase function. Although often not required, we restrict the maximum error by approximation to 1% *which is a very low tolerance with respect to rather moderate sample numbers. This way, we evaluate the conditional failure probability function* q *even with respect to utterly small failure probabilities which might not be relevant but avoid us to base results in the forthcoming simulation study on approximation errors. Higher admissible errors will often be sufficient and result in lower computational demands due to reduced total subset number* m^* *and increased SuS-like steps* m_s*.*

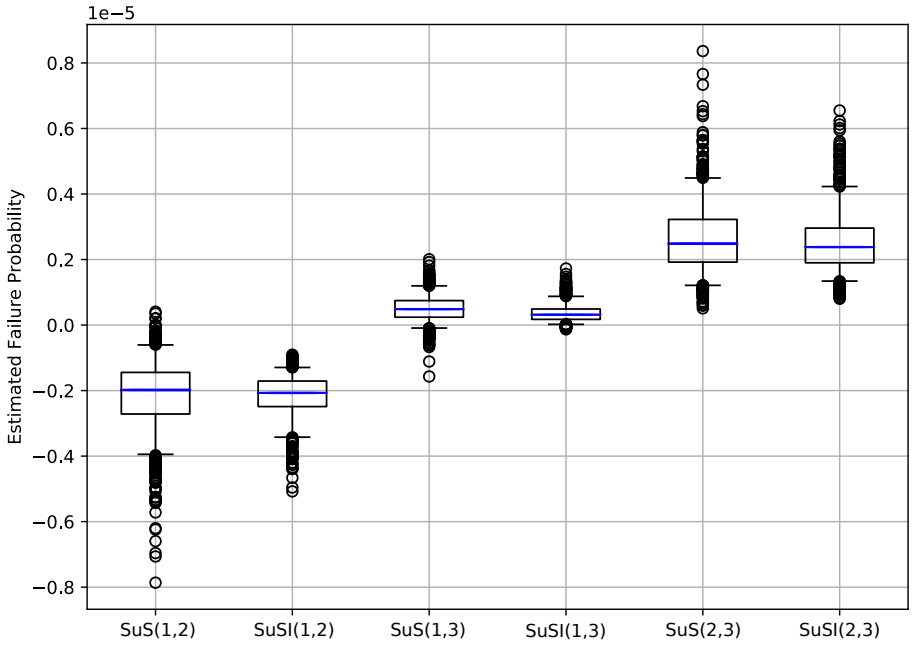
Remark 6.1.1, Remark 6.2.2, Remark 6.1.5 and Remark 6.1.6 underline that it is not straightforward to compare efficiencies of SuS and SuSI. Indeed, one should rather interpret these comparative results with respect to SuS, given in the following section, as an easy understandable way to demonstrate the computational demands of the introduced novel algorithm by relating it to the demands of a popular algorithm for a static evaluation.

6.1.1 Results

To study results by simulation, we have to distinguish between requiring the prediction step and the opportunity to use previous results for selection of adequate grid point values. We start with version a) of the prediction step, which in contrast to version b), explained later, is always available.

Prediction Step: Version a) First, we examine a typical form of a use-case of SuSI, given a dynamic model where the failure probability of a structure has to be estimated with respect to several different constellations. For illustration, we use g_1 with dimension two (for higher relevance of $\mathbf{X}_k$) and identify one of the variables as $\mathbf{X}_k$ and vary its distribution. Then, the estimated differences of the results are examined. Figure 6.3 yields the result. From this example, we may already conclude that SuSI might not only provide lower computational demands for several constellations, but also allows us to prevent drawing conclusions out of stochastic noise if differences of failure probabilities with respect to different constellations are studied. We now know that SuSI can perform very well in specific settings. Although SuSI uses about a third of the SuS samples, it is even better in estimation, avoiding to consider stochastic noise of single outcomes when comparing their estimated failure probabilities. In particular, in cases (1,2) and (1,3) we see that SuSI

Figure 6.3: Evaluating the difference in failure probabilities of three different constellations by SuS and SuSI. In every case, we rely on the limit state function g_1 with two variables. The first is fixed standard Gaussian always and the second one is varied to (1) standard Gaussian, (2) Gaussian with coefficient of variation 1.1 and (3) Gumbel with coefficient of variation 0.5. SuS(1,2) corresponds to comparing the results of 1000 Subset Simulation runs of constellation (1) and constellation (2), subtracting results by (2) from the ones by (1). Correspondingly SuSI(1,2) refers to the result by SuSI with similar number of total evaluations E_T for all three evaluations as for one simulation run by SuS.



does always provide the right sign where on the other hand SuS may not, so that its results are more reliable where the corresponding approximations by SuS might provide totally wrong results, e.g. if used for Taylor approximations afterwards. This may prevent extreme misinterpretations and therefore possibly severe consequences. However, there are some novelties such as the prediction step and the estimation by integration over estimated conditional failure probabilities which might affect its performance in an unexpected and possibly unfavorable manner. Furthermore, SuSI allows for several parameter settings that might lead to different conclusions. Thus, we need a more broad study.

As a first more extensive study, we fix the failure probability to $p_f = 1 \cdot 10^{-6}$ and track mean value, rRMSE and MAE for limit state functions g_1, g_{2a} and g_{2b} with respect to different parameter choices in SuSI. The result is shown in Table 6.1. Table 6.1 gives a good overview of several effects and illustrates how different choices of $[p_l, p_u]$ affect the distance of lower and upper staircase values to the real failure probability p_f. We see that moderate choices of $[p_l, p_u]$ appear to be favorable over rather extreme ones. As a consequence and by further experimenting, we choose $[p_l, p_u] = [0.2, 0.4]$ for further analysis. Note that the intervals were chosen to be rather large, since in Chapter 4 we have shown that efficiency does not significantly depend on the exact choice of the intermediate probabilities if $p_0 \geq 0.1$ in implementation (I4*) and to have higher probabilities for candidate acceptance in the prediction step, lowering the total number of necessary evaluations. An overview on the resulting number of evaluations is given in Table 6.2, showing that the samples for the prediction step took around half of the number of total samples in SuSI for dominant $\mathbf{X}_k$. This is due to the generally small number of samples in these examples and that we did not spent much efforts to reduce the number of samples used in the prediction step (e.g. by sophisticated adaptive methods). Also note that the number of total evaluations with prediction step was chosen slightly higher than for SuS because for less relevant $\mathbf{X}_k$ it decreases (compare Table 6.3, due to the increasing number of initial SuS steps) and we will later also consider variant b) where there are no prediction step samples, allowing to directly compare results of version a) with version b) (explained later) of the prediction step while having the same number of samples for SuS as a comparison. Typically, one would start with a single evaluation by version a) and then add some (e.g. $5 - 20$) simulation runs with version b) based on the first simulation result. Indeed we do not have to be very accurate with respect to the number of total evaluations

Scenario	**Method**	SuS	SuSI [staircase]		
M-Nm	N^*	1500	≈ 650		
$d = 2$	p_0	0.1	0.5		
$p_f = 1 \cdot 10^{-6}$	$[p_l, p_u]$	-	$[0.05, 0.15]$	$[0.2, 0.4]$	$[0.4, 0.6]$
$E_T \approx 10,000$	N^{pred}	-	100	75	50
$\hat{\mathbf{p}}_\mathbf{f} \cdot 10^{-6}$	g_1	1.03	1.25 [0.43, 8.49]	1.04 [0.67, 1.82]	1.14 [0.82, 1.62]
	g_{2a}		4.10 [0.52, 63.1]	1.32 [0.69, 6.10]	1.89 [0.85, 6.10]
	g_{2b}		1.07 [0.23, 1.90]	1.09 [0.64, 2.56]	1.11 [0.80, 1.37]
rRMSE	g_1	0.35	0.61 [0.61, 17.3]	0.40 [0.42, 1.09]	0.52 [0.40, 0.95]
	g_{2a}		15.7 [1.39, 345]	5.22 [0.70, 65.1]	19.9 [0.46, 118]
	g_{2b}		0.44 [0.78, 1.19]	0.48 [0.46, 32.4]	0.48 [0.40, 0.70]
MAE	g_1	0.26	0.42 [0.58, 7.48]	0.30 [0.37, 0.84]	0.37 [0.34, 0.69]
	g_{2a}		3.19 [0.70, 62.1]	0.54 [0.42, 5.12]	1.10 [0.37, 5.13]
	g_{2b}		0.33 [0.78, 0.92]	0.34 [0.41, 1.63]	0.36 [0.34, 0.50]

Table 6.1: **Efficiency of SuSI with respect to different limit states and parameter choices.** Results are based on 1000 independent simulation runs for SuSI and 2000 for SuS. The total amount of evaluations E_T under re-use of samples is approximately kept constant for all the choices, resulting in different N with respect to the used method and parameters. Underlying limit state function and basic variable distributions are as in Papaioannou et al. (2015) Example 1 (linear), Example 2a (convex) and Example 2b (concave) which we already used in the simulations of Chapter 4. Implementation (I4*) is used. N^* relates to $p_0 = 0.1$ in every case, where for the real selected N in subsets, we use $N = N^* \cdot (\log(p)/\log(0.1))$ for $p = p_0$ or $p \in [p_l, p_u]$. Except for limit state function g_{2a}, results by interpolation with splines appear to be slightly higher, but similar, as those given by SuS for $[p_l, p_u] = [0.2, 0.4]$.

	$E_T - \sum N^{pred}$			$\sum N^{pred}$			E_T		
$[p_l, p_u]$	g_1	g_{2a}	g_{2b}	g_1	g_{2a}	g_{2b}	g_1	g_{2a}	g_{2b}
$[0.05, 0.15]$	10500	9500	8500	3000	4000	5500	13000	13500	14000
$[0.2, 0.4]$	8500	8000	7500	3000	4000	3500	12000	12000	10500
$[0.4, 0.6]$	8000	7500	7000	5500	5500	4000	13500	13000	11000
SuS	9000	9000	9000	0	0	0	9000	9000	9000

Table 6.2: **Number of evaluations with respect to the simulations in Table 6.1** (Dimension $d = 2$ of the limit state functions). The total number of evaluations for SuS and SuSI as well as its split into evaluations by the prediction step and normal subset evaluations are provided. $\sum N^{pred}$ corresponds to the sum of all prediction step evaluations under consideration of all subset levels.

	$E_T - \sum N^{pred}$			$\sum N^{pred}$			E_T		
$[p_l, p_u]$	g_1	g_{2a}	g_{2b}	g_1	g_{2a}	g_{2b}	g_1	g_{2a}	g_{2b}
$[0.05, 0.15]$	8500	8500	6500	2000	3500	1500	10500	12000	8000
$[0.2, 0.4]$	7000	7000	5500	2500	3500	2500	9500	10500	7000
$[0.4, 0.6]$	6500	6500	5500	4000	5000	2500	10500	11500	8000
SuS	9000	9000	9000	0	0	0	9000	9000	9000

Table 6.3: **Number of evaluations with respect to limit state functions with dimension d = 10.** As in Table 6.2, the total number of evaluations for SuS and SuSI are provided.

here, as our focus is on general properties and proof of concept. Regarding the choice of the parameters in SuSI, we found $[p_l, p_u] = [0.2, 0.4]$ to perform best. Additionally, we see bad results with respect to limit state function g_{2a} for all choices of $[p_l, p_u]$. For a better understanding, an analysis with respect to the quantiles is considered in Figure 6.4 and Figure 6.5. A closer look at the quantiles exposes that the bad rRMSE must be the consequence of rare events. Parameters $[p_l, p_u] = [0.2, 0.4]$ were also found favorable in analysis of quantiles, although less than expected by Table 6.1. The following remark explains our findings and the suggestion to favor moderate values of $[p_l, p_u]$.

Remark 6.1.7 (Discrepancies: Stochastic Prediction Step a))**.** *In Chapter 4, we have shown insensitivity of the coefficient of variation of the results by SuS with respect to SuSI. By Section 5.3 and as we just integrate over several SuS results in the same manner always, the key to differences in results with respect to the parameter settings, beyond the effect on the total number of evaluations, must be assigned to the prediction steps. Indeed, the*

Figure 6.4: Analysis of the quantiles of estimates by SuS and SuSI, based on 1000 simulation runs in each setting. The results refer to an important dynamic variable, setting the dimension of limit state functions g_1, g_{2a} and g_{2b} to two ($d = 2$).

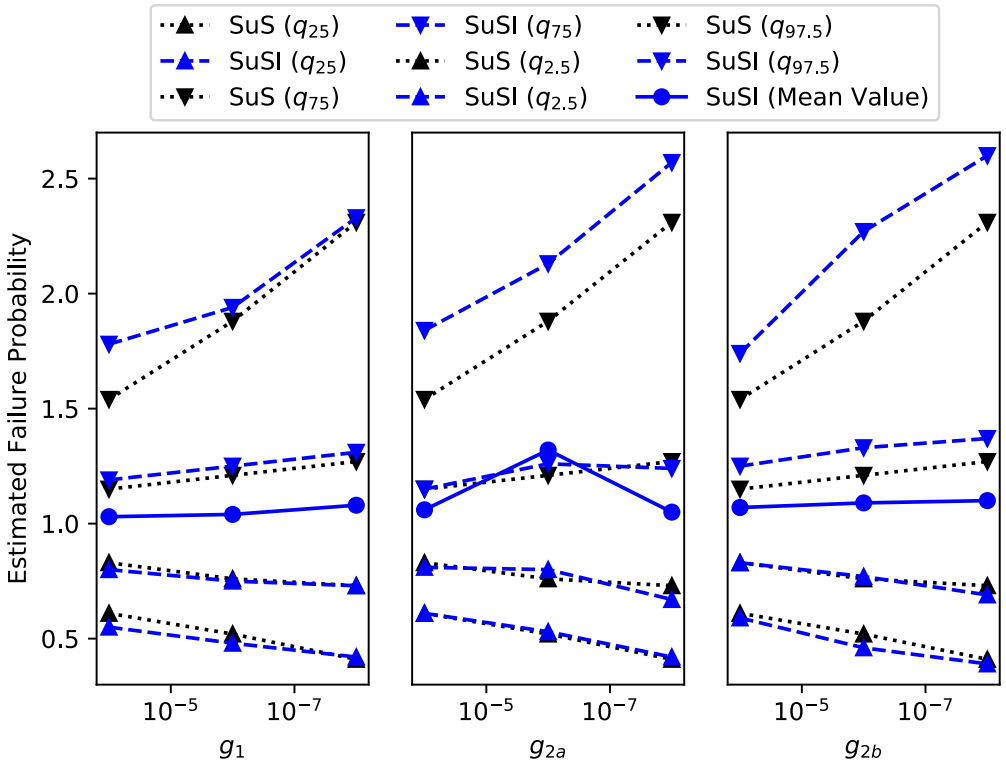
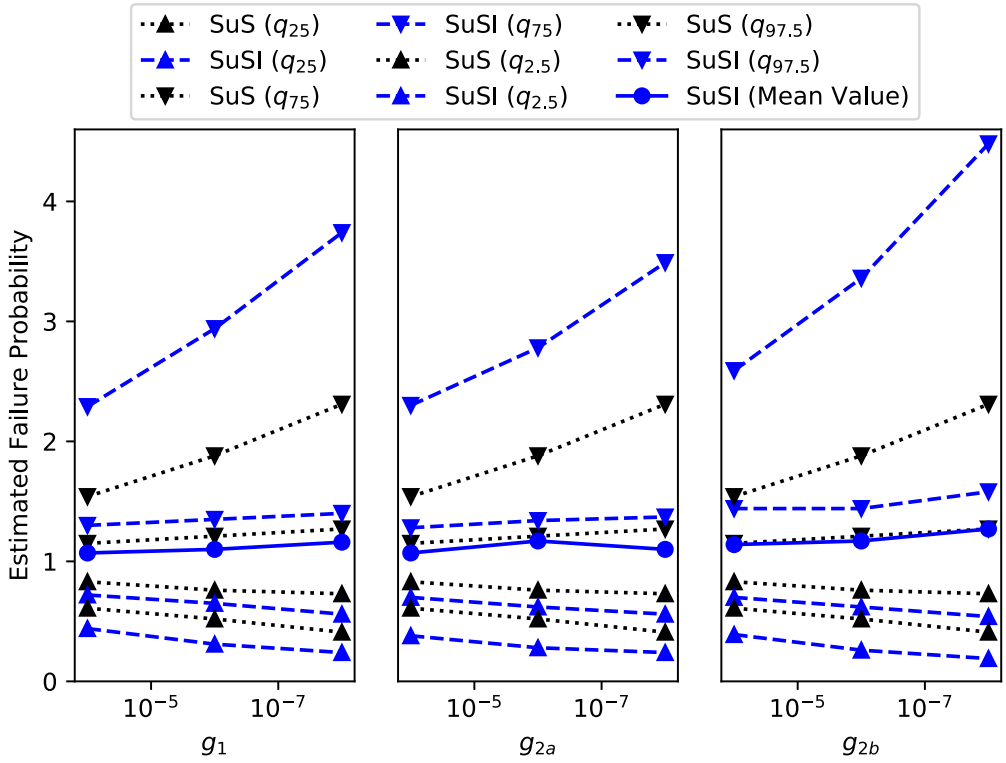


Figure 6.5: Analysis of the quantiles of estimates by SuS and SuSI, based on 1000 simulation runs in each setting. The results refer to a rather unimportant dynamic variable, setting the dimension of the limit state functions g_1, g_{2a} and g_{2b} to ten ($d = 10$).



shortcomings for $[p_l, p_u] = [0.05, 0.15]$ *and* $[p_l, p_u] = [0.4, 0.6]$ *can be explained by the impact of this selection on the prediction steps. Scanning for very low intermediate probabilities will result in high effects for stochastic variation in the predictions. Selecting slightly wrong grid points might result in an extreme relative change of the intermediate probability, yielding to* $p_0 < 0.05$ *which will lead to a bad performance. Choosing* $[p_l, p_u] = [0.4, 0.6]$ *on the other hand requires to lower the number of samples for the prediction step if we do not want to increase the number of total evaluations much and also the requirement for several prediction steps because the number of subsets increases a lot. As a result, we more frequently have bad results due to the stochastic variation of the result by the prediction step. In general, estimating an intermediate probability by a prediction step with less samples than used for evaluation of the subsets, will usually lead to a non-optimal choice of the number of subset samples in the corresponding intermediate probability evaluation, since this has to be set in advance. However, such methods could also be made adaptive again, increasing the performance.*

Remark 6.1.7 discusses choices of $[p_l, p_u]$ and also explains the shortcoming of a stochastic prediction step. In particular, we found high rRMSE and high mean values to be the consequence of some rare events, where the prediction step did not provide sufficiently well estimated grid point values.

On the one hand, these rare events would be a big drawback, causing the algorithm to perform bad on average although yielding a good efficiency most of the time. On the other hand, we have some options to avoid them:

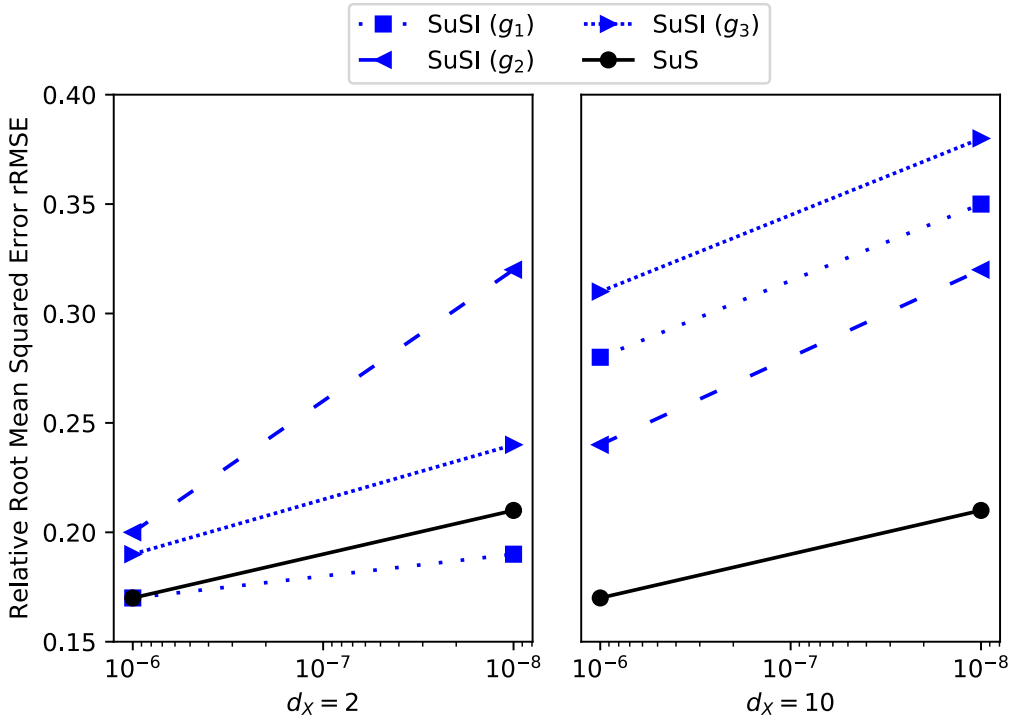
- Generally, it was easy to identify ill-posed results. Typically intermediate probabilities take extreme values, either close to zero (e.g. below 0.01) or approximately one, in the rare events where the prediction step fails. As these are visible, such results can be detected and removed in a post-processing step.

- We can choose to use more samples for the prediction step.

- Version b) of the prediction step can be utilized after we have one terminated simulation. Typically, we should make several independent simulation runs, as in Subset Simulation, to evaluate results, anyways.

- The prediction step can be further refined. Many adaptive methods can be tested.

Figure 6.6: Comparison of the rRMSE of SuS and SuSI. The dimension of the limit state function is set to two (left) and ten (right). High sample numbers are considered.



Using more samples in the prediction step is always possible and easy, but results in a higher number of total evaluations. Nevertheless, for generally higher sample numbers, one could naturally increase the samples for the prediction step adequately. This further increases robustness of the prediction step. The result for doubled prediction step sample number and five times the sample numbers considered in Table 6.1 is given in Figure 6.6. Under doubled sample number in the prediction step, no ill-posed rare events were present anymore, in 1000 independent simulations. According to Figure 6.6, we see a similar, although often slightly higher, rRMSE for dimension $d = 2$. In the case of dimension $d = 10$, SuSI clearly has a worse rRMSE. This was expected and we also point out again that the most interesting cases are the ones where the dynamic variable is more relevant such as for $d = 2$.

Thus, not only because of the intended application on dynamic models, but also because of this fact, we claim the algorithm to be very efficient. Also note that our expected main drawback, the bias, seems to be negligible in many cases, as we did not have systematic errors in our simulations. Results with high N, corresponding to Figure 6.6, were found to not provide a significant bias, taking its maximum at $d = 10$ and g_{2b} with 5%. While there exist functions that provide the maximum bias derived by statistical analysis in Section 5.3, these functions seem to require an unnatural shape, at least for moderate $[p_l, p_u]$ values such as $[0.2, 0.4]$. As a consequence, we focus on interpolation by the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP), only rarely examining staircase approximations.

If several independent simulation runs of SuSI are used for reliability evaluation, then we should use previous results for more accurate prediction steps ('version b' of the prediction step). Although the algorithm already performs similar as SuS in many cases of static settings for high N, the presented algorithm is a novel approach so that many enhancements may still be possible with respect to its implementation.

Prediction Step: Version b) Next, we consider version b) of the prediction step so that the simulation procedure is as follows. The first simulation run is executed under prediction step version a). Then, all successive runs are carried out under selection of the next grid point by using the previous estimated conditional failure probability function $\hat{q}$ with respect to the existing simulation results, version b) of the prediction step. This does not result in correlated results, since this procedure only has an effect on the grid point selection, not on the estimation of intermediate probabilities with respect to the stochastic variation. Variant b) does not only decrease the sample number in simulations, but at the same time might increase the goodness of fit of the grid point estimators as previous results are based on the usually higher sample number per subset instead of the sample number for prediction steps N^{pred}. For an even higher benefit, one can also use several results of terminated simulations for estimation in the prediction step. The result is shown in Figure 6.7. As expected, we furthermore avoid the outliers, rarely available but strongly influencing the rRMSE and the mean value in version a) so that we can now consider the rRMSE as a performance measure instead. Table 6.2 and Table 6.3 show that the number of samples is reduced significantly in version b), neglecting prediction step samples. In comparison

Figure 6.7: Comparison of the rRMSE of SuS and SuSI, using version b) of the prediction step in SuSI. The dimension of the limit state functions is set to two (left) and ten (right).

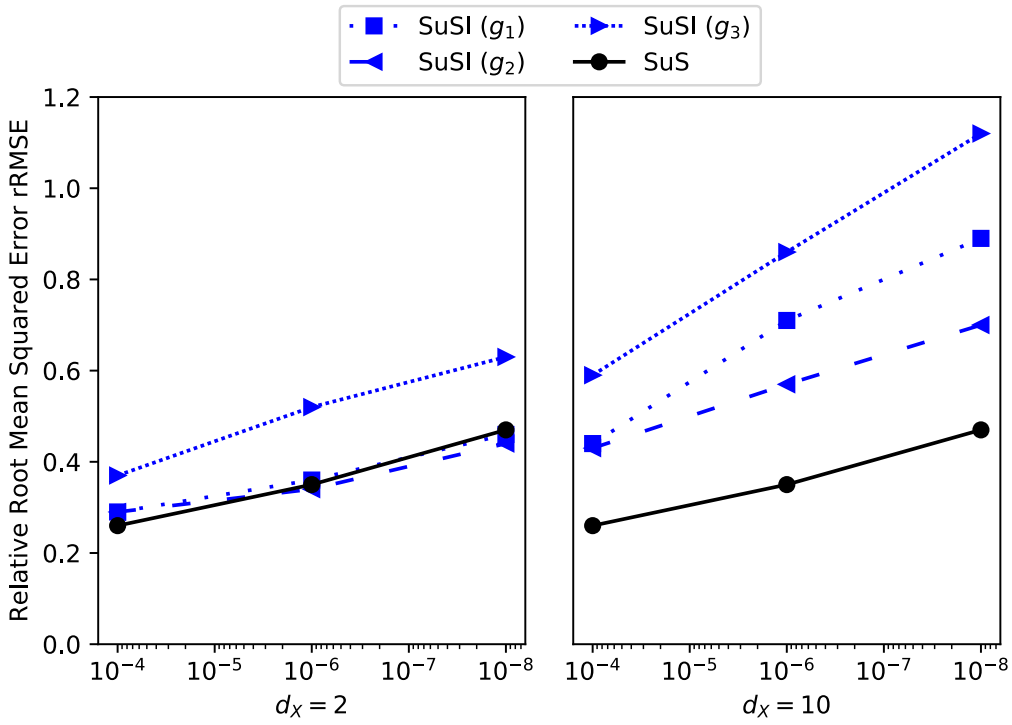
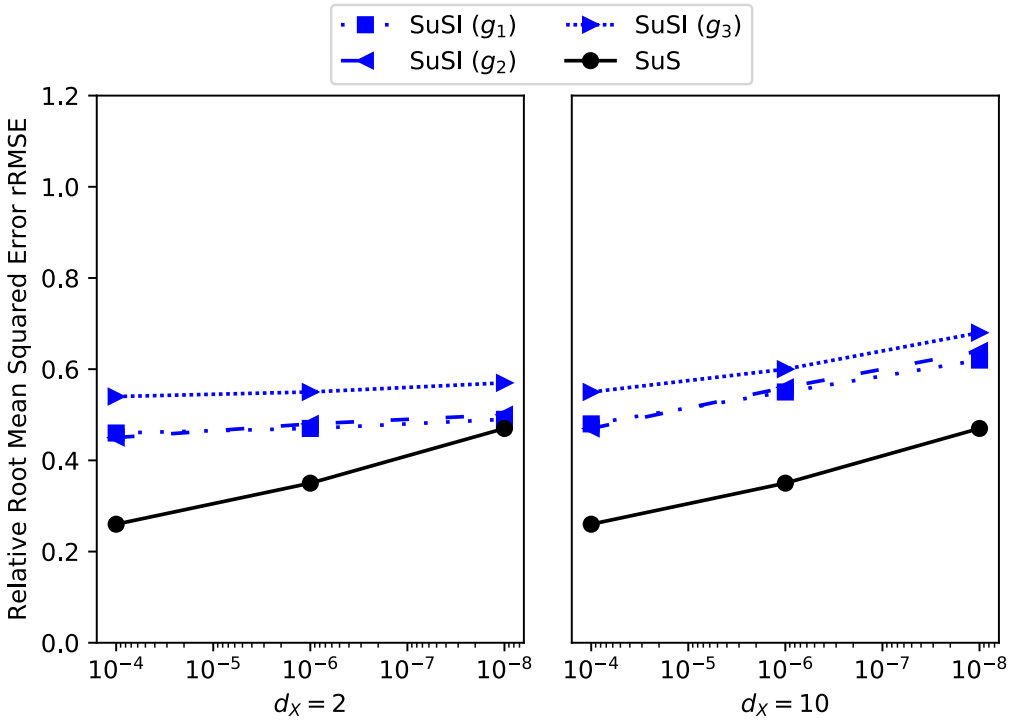


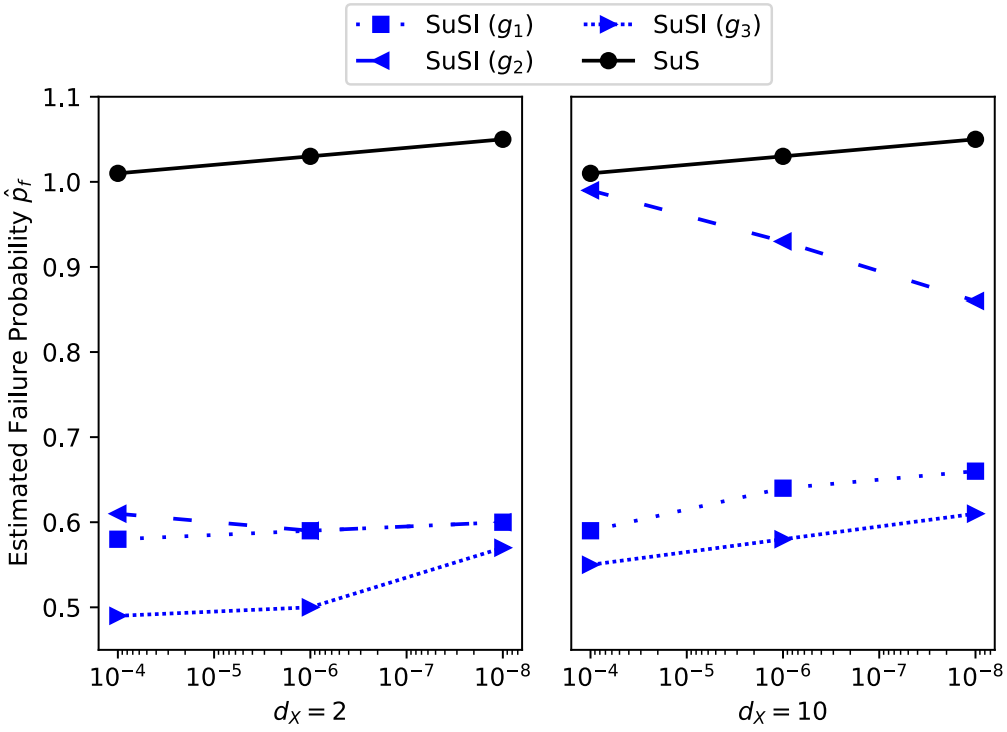
Figure 6.8: Comparison of the rRMSE of SuS and SuSI under usage of lower staircase approximation in SuSI. The dimension of the limit state functions is set to two (left) and ten (right).



to SuS, we then even have a smaller number of total evaluations E_T in SuSI. Although using $[p_l, p_u] = [0.2, 0.4]$ for examination, simulations suggest that prediction of the adaptive grid point selection by previous simulation runs allows to switch to higher intermediate probability values for more assurance of a low bias, also. Still, for ease of comparison and also because the bias seems to be negligible in most cases, we keep $[p_l, p_u] = [0.2, 0.4]$ fixed in the analysis.

As already supposed in the statistical analysis part, it is sometimes possible to use the lower staircase approximation to decrease the rRMSE and thus increase efficiency. An example is given in Figure 6.8. In comparison to the results by interpolation (compare Figure 6.7), particularly in the case of less relevant $\mathbf{X}_k$ (respectively $d = 10$) the rRMSE is drastically reduced.

Figure 6.9: Comparison of the empirical mean values of the estimated failure probabilities under consideration of the lower staircase approximation in SuSI. The dimension of the limit state functions is set to two (left) and ten (right).



Also, note that the shape of the conditional failure probability function seems to lose relevance and better results could even be achieved for lower sample numbers so that the bias has less impact on the total result. However, we then underestimate the true failure probability systematically which requires to be extra careful. The corresponding mean values are shown in Figure 6.9.

As a last simulation study, we consider a more practical example, a two degree of freedom primary/secondary damped oscillator (see Der Kiureghian and De Stefano (1991); Bourinet et al. (2011)). We fix the mean value of the force capacity of the secondary spring F_s to $\mu_{F_s} = 27.5$, resulting in a reference failure probability of $p_f = 3.78 \cdot 10^{-7}$. The limit state function is

Variable	m_p	m_s	k_p	k_s	ζ_p	ζ_s	F_s	S_0
Distribution	ln	ln	ln	ln	ln	ln	ln	ln
Mean	1.5	0.01	1	0.01	0.05	0.02	27.5	100
c.o.v.	0.1	0.1	0.2	0.2	0.4	0.5	0.1	0.1

Table 6.4: Bourinet et al. (2011). Random variables according to limit state function (6.1) by Der Kiureghian and De Stefano (1991).

given by (Der Kiureghian and De Stefano (1991), Bourinet et al. (2011)):

$$g(x) = F_s - 3k_s\sqrt{\frac{\pi S_0}{4\zeta_s\omega_s^3}\left[\frac{\zeta_a\zeta_s}{\zeta_p\zeta_s(4\zeta_a^2+\theta^2)+\gamma\zeta_a^2}\frac{(\zeta\omega_p^3+\zeta_s\omega_s^3)\omega_p}{4\zeta_a\omega_a^4}\right]} \tag{6.1}$$

with $\omega_p = \sqrt{k_p/m_p}$, $\omega_s = \sqrt{k_s/m_s}$, $w_a = (w_p + w_s)/2$, $\zeta_a = (\zeta_p + \zeta_s)/2$, $\gamma = m_s/m_p$, $\theta = (\omega_p - \omega_s)/\omega_a$ and random variables according to Table 6.4, given distribution type, mean value and coefficient of variation (c.o.v.) of each variable. We consider four cases for SuSI, successively setting ζ_p, ζ_s, F_s and S_0 as the dynamic variable $\mathbf{X}_k$. For analysis of efficiency, we use 50 independent simulation runs, each starting with a single estimation according to SuSI with prediction step variant a) followed by 19 simulation runs with prediction step variant b). In variant b), we consider only the previous simulation run for estimation of suitable grid points. In total, we thereby get 1000 simulation runs. For comparison of performance, we use 2000 independent simulation runs of SuS. The resulting statistics of the results by simulation are described in Figure 6.10 and Figure 6.11. The number of repetitive runs for testing in this example is small. However, for general statements these should be sufficient at this point, in particular because the results based on averages of repetitive runs make the corresponding distribution of the results less heavy tailed. SuSI with dynamic variables ζ_p, ζ_s and F_s appears to provide a similar performance as SuS does. Choosing S_0 as a dynamic variable, on the other hand, requires the computational demands of a few static reliability analysis by SuS, with same number of total evaluations in each analysis, to achieve the same accuracy. Remarkable is that the total number of evaluations in SuSI was smaller than it was for SuS on average, still providing similar efficiency as SuS for several choices of dynamic variables so that we expect a practically equivalent coefficient of variation for both methods under same number of total evaluations in this static evaluation. Indeed we took the same parameter settings as in Table 6.1 with $[p_l, p_u] = [0.2, 0.4]$. The average samples in the

Figure 6.10: Box plots of the estimated failure probabilities of SuS and SuSI, based on 1000 simulation runs. Whiskers correspond to 5 and 95 percent quantiles. SuSI was applied with respect to variables ζ_p, ζ_s, F_s and S_0.

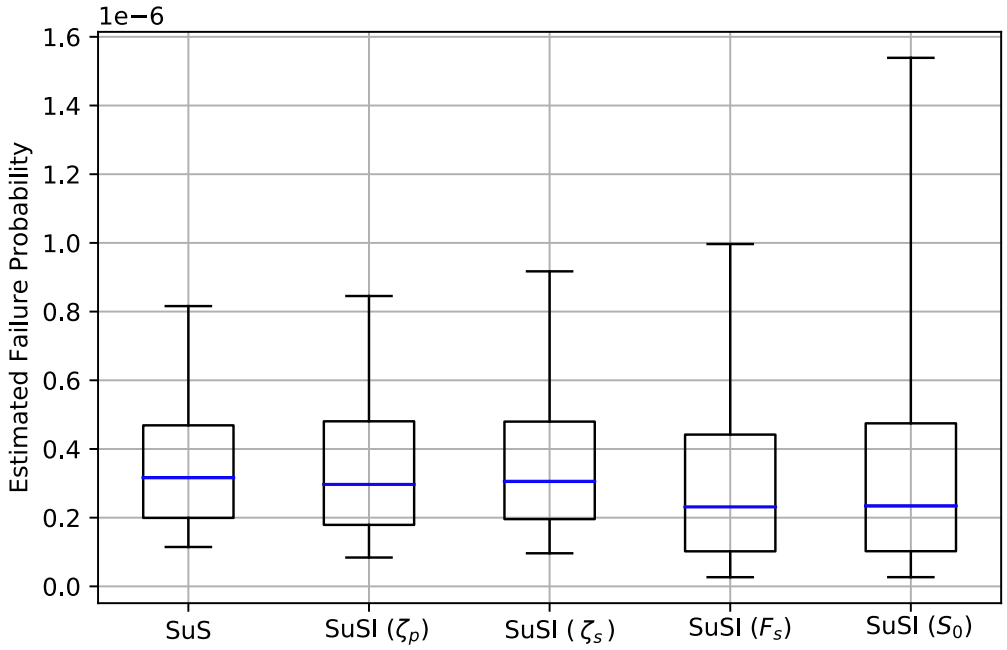
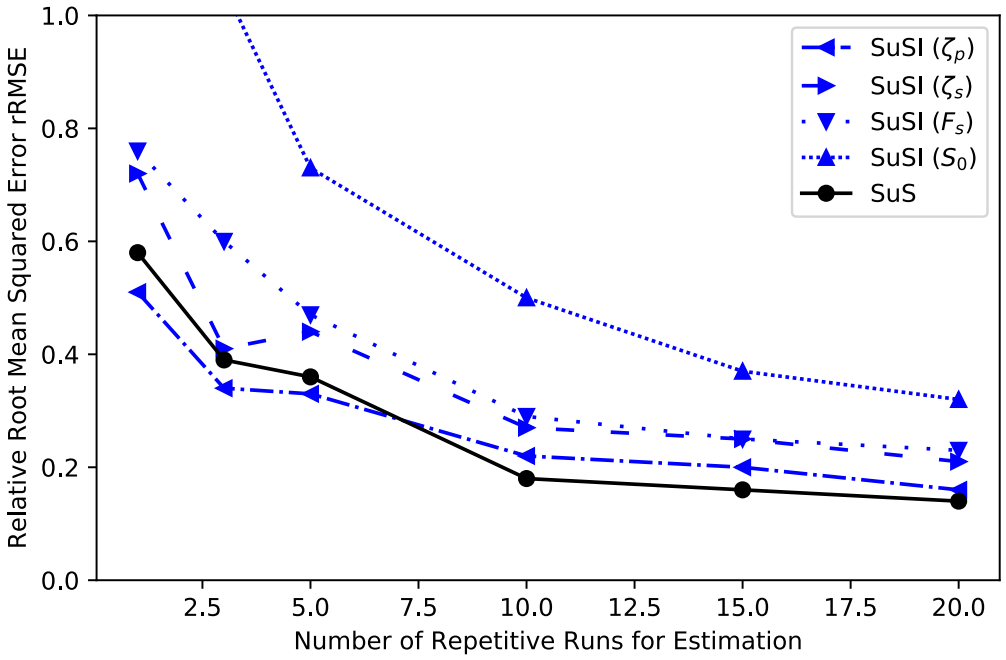


Figure 6.11: Usage of repetitive calculations for estimation. Based on 50 independent simulation runs, each consisting of a first simulation run with prediction step a), followed by nineteen simulations with prediction step b), we plot the change of rRMSE with respect to utilization of an increased number of repetitive runs. SuSI was considered with respect to four different stochastic variables. For comparison, 100 SuS runs were considered in the same manner but without any variation in parameter settings.



prediction step become negligible for higher (say more than 5) number of repetitive runs. So, we get the result for a dynamic model at the cost of a static reliability evaluation by SuS in a realistic setting. Also note that the possibility to condition on several variables allows to draw conclusions based on this joint information, while the failure probability of the static reliability problem is simultaneously computed by several simulation runs. We can, for example, compare sensitivities with respect to different variables or provide evidence for accurateness of the results in contrast to calculation by SuS where reliability evaluation relies on the same settings for the limit state and the stochastic variables in every repetitive run.

Lastly, we want to remark that we did not focus on optimizing every single parameter of SuSI for testing the new algorithm since our focus was on introducing the general concept and demonstrating functionality of it. Indeed, we think that there exist many opportunities for improvement. Some of them were shortly discussed previously, some are left for future research.

6.1.2 Conclusions

Concluding, our simulations give evidence for functionality of the method and support our findings by statistical analysis. We have two main conclusions.

First, it is remarkable that the bias by interpolation, which was first identified as the potential drawback of the algorithm, was found to be small in the considered problems and there even exist several options to easily reduce it further such as refinement of the grid as a post-processing procedure.

- In the simulation studies, the bias was found negligible, if systematic errors induced by the stochastic prediction step were excluded by proper implementation or post-processing.
- There are several options for a certain bias reduction:
  - If repetitive runs of the algorithm are performed, regression can be used.
  - The set of grid points can be refined in an additional post-processing step. Such additional evaluations can focus on important regions of the domain of $\mathbf{X}_k$ and even require a lower computational effort than usual evaluations, because the samples for these evaluations are already given and it is also known, by monotonicity, that some samples surely fail according to the successive subset.

– Intermediate probabilities can be increased.

The regression approach was not studied in this work, since we first focus on building a strong basis for the generally better traceable interpolation case, which is also applicable after a single run of the algorithm. To study regression in detail, many settings would have to be accounted for, such as the number of repetitive runs considered, the number of selected knots and the degree of smoothness. This would go beyond the scope of this work, which focuses on introduction of the algorithm and its general applicability, and is therefore left for future research.

Second and most importantly, the algorithm yields an impressive performance as it has often similar, although sometimes slightly higher, computational demands as a static reliability evaluation by SuS for analysis of a much more complex dynamic reliability model. Therefore, SuSI can reduce the necessary computational effort in dynamic reliability evaluation, drastically. On the other hand, there exist cases where SuSI requires a significantly higher computational demand to achieve similar accuracy as SuS in one static reliability evaluation, such as an irrelevant $\mathbf{X}_k$. This however is not necessarily a big disadvantage, since good performance was shown in the use-case.

Remark 6.1.8 (Use Case: Similar Computational Demands as a Static SuS Evaluation)**.** *As SuSI is designed for dynamic models, it should be applied when* $\mathbf{X}_k$ *is expected to have a rather high impact on the failure probability and particularly when differences between different settings have to be considered. In both cases, SuSI achieves at least a similar* rRMSE *as SuS with the same number of total evaluations in a static reliability evaluation (compare Figure 6.3, Figure 6.6 (left) and Figure 6.7 (left)). Worthwhile is in particular the superior performance of SuSI in computing relations of the failure probabilities of different model constellations by avoiding stochastic noise in the results, as shown in Figure 6.3.*

In contrast to Remark 6.1.8, in cases where $\mathbf{X}_k$ is not so relevant, SuSI might need up to about five times as many evaluations as SuS in a corresponding static reliability evaluation, in our examples. Nevertheless, this can still be considered an acceptable computational demand for the provided gain in information.

6.2 Application: Model Dynamics

Important to note is that we identify the main benefit of SuSI not as an increased efficiency given n constellations. This setting was useful for tractability of performance. Instead, SuSI allows a different way of approaching reliability analysis under model dynamics and specific uncertainties or robust analysis by accurate continuous failure probabilities with respect to the distribution of a stochastic variable.

New Methods allow to pose and answer new questions, helping to extend the understanding in structural reliability analysis.

The provided information by SuSI makes it straightforward to use completely different representations of the reliability with respect to variation in the distribution of $\mathbf{X}_k$. However this is not only beneficial for specific distributions, it also gives new insights such as the conditional failure probability with respect to deterministic values of $\mathbf{X}_k$, even if it has a very complex and not at all comprehended impact on the failure probability in advance. So, extensive case studies or analysis that include many different distributions of a variable are easy, cheap and handy to include into analysis. As a consequence, it becomes more likely to discover effects that remain unseen otherwise. Furthermore, there are many more opportunities. For example the shortly presented robust evaluation (presented later, in Section 6.2.3) which could be very valuable for backup testing, or as a stand-alone algorithm, could only be one of many such opportunities. Especially because of the widespread application of SuS in many different fields with individual requirements for good and interesting aspects of analysis, we will most likely not even be able to demonstrate some obvious use-cases with a high benefit in specialized fields.

Anyhow, we present several cases for application of the algorithm, where for one variable $\mathbf{X}_k$, several constellations of the distribution function need to be evaluated. In particular, we demonstrate that $\mathbf{X}_k$ is not restricted to the original variables of the problem formulation, but can be anything, such as a time variable or even a variable that artificially changes the failure set drastically by scaling the whole failure space.

Remark 6.2.1 (Complex Structures)**.** *Note that the given examples for application in this section are for illustration purposes and therefore do not claim absolute correctness or relevance on the engineering side, nor are they*

complex enough to require applying SuS. Picking simple examples for illustration allows for an easy access to the settings of the examples and to the general understanding of the presented procedures. In Chapter 8, we explicitly address a few more complex problems, where SuSI could be applied.

However, the examples are still well suited for illustration and the presented procedures can be applied to complex structures and higher dimensions just as well and in the same manner.

Remark 6.2.2. *Also worthwhile, beyond the high number of cheap reliability evaluations, is that the algorithm might reduce stochastic noise when comparing several different constellations (compare Section 6.1). These relations can then be derived by the same stochastic result, instead of relying on independent results where the randomness of results can be misinterpreted as information on the relation of failure probabilities. Examples which benefit from this advantage include sensitivity analysis and time-dependent reliability analysis, presented in the following.*

Coefficients of variation of variables of the stochastic model are generally referred to as c.o.v., or written as e.g. v_{f_c} for stochastic property f_c, in contrast to the coefficient of variation $\mathrm{CV}(\hat{P}_f)$ of the estimator of the failure probability.

6.2.1 Extensive Sensitivity Analysis[2]

We start with a reinforced concrete beam, such as in Figure 6.12, where the underlying probabilistic model is given in Table 6.5. The model is based on Glowienka et al. (2011) with modified loads and an extra limit state. In detail, the beam is exposed to distributed loads $g + q$ and a test (or accidental) load F_a that is concentrated at one ending of the beam where we are interested in two failure mechanisms, reinforcement failure due to flexural tension ($F_{\mathbf{reinf.}}$) and failure of the concrete by crushing of the compression strut ($F_{\mathbf{conc.}}$). The corresponding limit state equations are given by:

$$\begin{aligned} g_{\mathbf{reinf.}} =& \theta_{R,M}(f_y A_s)\left((h-d_1) - \frac{f_y A_s}{1.615 b f_c}\right) \\ & - \theta_{E,M}(M_{g+q} + M_a) \end{aligned}$$

[2]This example was first presented and is extracted from Blandfort et al. (2019a).

Table 6.5: Blandfort et al. (2019a), modified. Probabilistic model for reinforced concrete beam example. Relevant distributions for the random variables are normal(n) and lognormal(ln). Additionally, there are deterministic(d) model variables. Units: Forces are given in meganewton[MN], geometric quantities in meters[m].

Variable	f_y	A_s	b	h	d_1	d
Distribution	ln	det	n	n	n	n
Mean μ	580	0.00493	1.0	0.8	0.05	0.75
c.o.v. v	0.06	-	0.01	0.013	0.2	0.013
Variable	b_w	α_c	$cot\theta$	a_1	a_2	
Distribution	n	det	det	det	det	
Mean μ	1.0	0.75	3.0	0.3	6.20	
c.o.v. v	0.01	-	-	-	-	
Variable	$\theta_{[-]}$	f_c	$g+q$	F_a		
Distribution	n	ln	n	det		
Mean μ	1.0	20-50	0.1	2.0		
c.o.v. v	0.1	0.1-0.2	0.2	-		

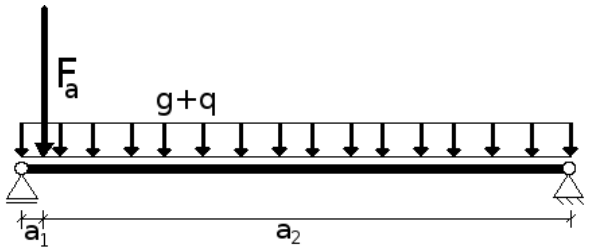


Figure 6.12: Blandfort et al. (2019a). Reinforced concrete beam exposed to distributed load $g + q$ and concentrated load F_a.

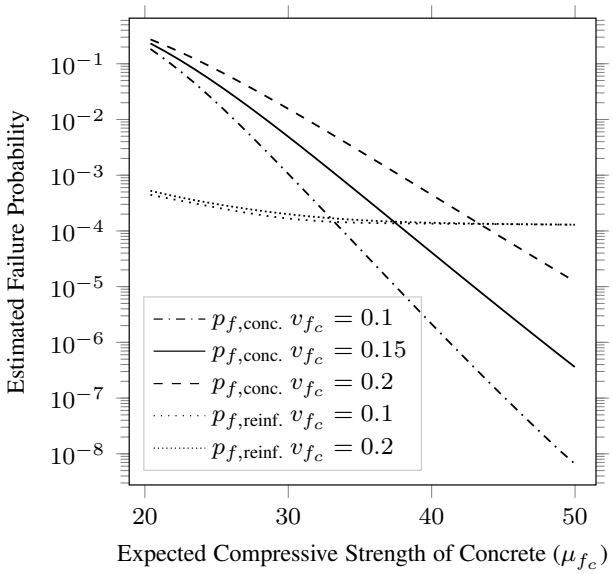


Figure 6.13: Blandfort et al. (2019a). Sensitivity analysis of failure probability and failure mechanisms of a RC beam with respect to its compressive strength distribution.

and

$$g_{\textbf{conc.}} = \theta_{R,V} \frac{b_w 0.9 d \alpha_c 0.85 f_c}{cot\theta + tan\theta} - \theta_{E,V}(V_{g+q} + V_a) \ .$$

Now, $g_{\textbf{reinf.}}$ belongs to a ductile, whereas $g_{\textbf{conc.}}$ belongs to a brittle failure mechanism. Thus, it is crucially important which failure mechanism is more likely to occur. Unfortunately, concrete core testing typically comes with huge uncertainties in estimation of the compressive strength f_c. For that reason, we examine reliability and dominant failure mechanism for several stochastic distributions of f_c. Applying SuSI, this only requires one single run for each of the limit state equations, then a one dimensional integration yields the result for each chosen distribution. Results are illustrated in Figure 6.13. For comparison, SuS would have required 150 simulation runs to compute failure probabilities for 5 different settings at 30 nodes. The result shows, as expected, a high dependency between failure due to crushing of the compression strut and f_c, whereas the impact on reinforcement failure due to flexural tension is insignificant. As a consequence, the dominant failure mechanism highly depends on the assumed distribution of f_c here. Furthermore, high variability of the failure probability with respect to the stochastic

distribution of f_c and the uncertainty involved when estimating it, may also lead to risks such as dangerous overestimations or costly underestimations of reliability. For that reasons, a rigorous analysis, including uncertainties, yields important information.

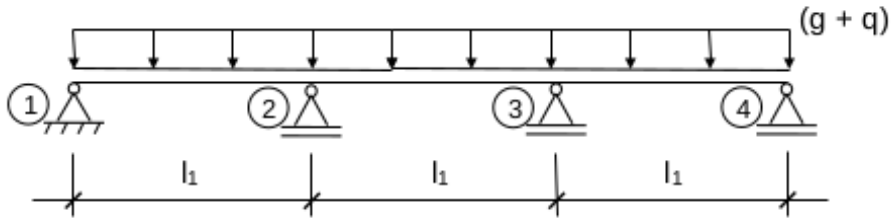
6.2.2 Time-Dependent Reliability Analysis

Here, we show how time-dependent models may very well suit the dynamic model in SuSI. Instead of choosing the dynamic variable $\mathbf{X}_k$ as a stochastic property of the structure, this time $\mathbf{X}_k$ will take the role of the time variable t. For discussion, we take a simple RC slab example, following Blandfort et al. (2021) but taking a different time-dependent approach in this section. U_M represents the model uncertainty of the resistance and is assumed to be lognormally distributed with mean value $m_S = 1.0$ and coefficient of variation $v_S = 0.07$. The limit state function for its dominant failure mechanism (bending) is given by

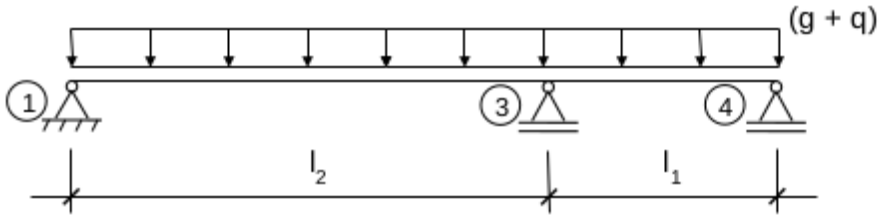
$$g(X) = U_M(f_y A_s)\left((h - d_1) - \frac{f_y A_s}{1.615 b f_c}\right) - |M_s|$$

where M_s represents the moment force in the column area 2. Loads are given as a combination of dead loads g and live loads q (assumed to act on all three fields at the same time). Dead loads are generally either given as deterministic or Gaussian with low variability. The stress resultant M_s is the resulting moment force derived by dead loads g and live loads q. We assume $g+q$ Gumbel distributed with mean value $m_{50} = 0.015\frac{MN}{m^2}$ and c.o.v. $v_{50} = 0.1$ for the reference period of 50 years. Then M_s is Gumbel distributed with mean $m_S = -0.1(g+q)l_1^2 = -0.02166MNm$ and coefficient of variation $v_S = 0.1$. The corresponding yearly maximum loading distribution, if no time effects are present, is then given by $(m_1, v_1) = (0.01505, 0.14388)$. An overview of the stochastic properties is given in Table 6.14c. Next we want to add time dynamics to the model and show that SuSI allows for an efficient evaluation. We consider two cases.

In the first case, we assume a variable of the structure that fulfills monotonicity (Assumption 2.2.2) and independence (Assumption 2.2.1) assumptions as dynamic variable that changes over time and has the most important effect on the failure probability of the structure. For example, the structure might be exposed to a marine environment and suffer from corrosion. The time-dependent decrease of the cross sectional area of the steel might then



(a) 3-span slab



(b) 2-span slab

Property	Values								
Variable	f_c	f_y	A_s	b	h	d_1	l_1	U_M	M_s
Distribution	ln	ln	d	n	n	n	d	ln	g
Mean m	33	580	0.000524	1.0	0.2	0.03	3.8	1.0	-0.01505
c.o.v. v	0.132	0.06	-	0.01	0.013	0.2	-	0.07	0.14388

$*$ f_c, f_y and M_s are given in $\frac{MN}{m^2}$, A_s in $\frac{m^2}{m}$ and the geometric quantities b, h, d_1, l_1 in m.

(c) Stochastic properties of the reinforced concrete slab

Figure 6.14: Blandfort et al. (2021). Properties of the reinforced concrete structure, 3-span slab (Figure 6.14a) and corresponding 2-span slab (Figure 6.14b) after column failure are given. Both are subjected to evenly distributed dead loads g and live loads q. The corresponding stochastic properties, except for the loadings, are given in Fig 6.14c.

have the highest impact on reliability over time. If, unfortunately, the exact cross sectional area and how it develops over time is mostly unknown but we, nevertheless, have to make a prediction on the remaining life time (or plan the next inspection) under sufficient reliability, we are required to consider several realistic scenarios for a meaningful conclusion. The procedure is then exactly the same as in the previous example, just that sensitivity analysis is now connected to specific time instants instead of relating to uncertainty considerations. However, a much more specific way to apply SuSI on time-dependent reliability analysis is the following.

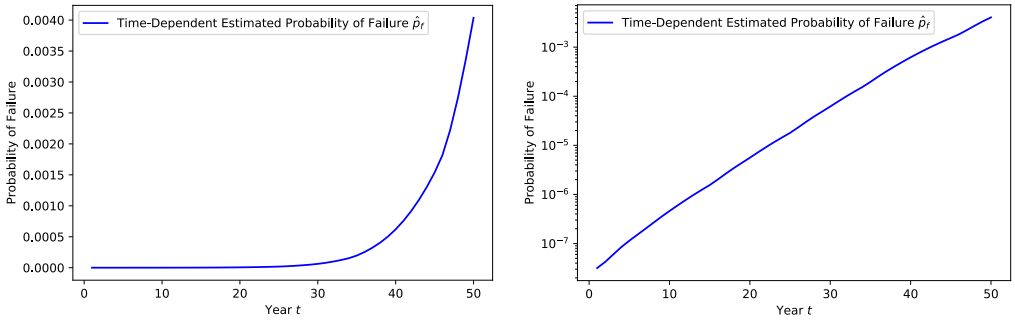
In the second case, we assume the time-dependent limit state function can be written with respect to t. For demonstration, it is assumed that the steel cross sectional area A_s decreases by $\kappa = 0.000002\frac{m^2}{m}$ per year and loading demands increase by 1% in every year. The resulting time-dependent limit state function is then given by

$$\begin{aligned} g(X,t) =& U_M(f_y(A_s - \mathbf{t}\cdot\kappa)\left((h-d_1) - \frac{f_y(A_s-\mathbf{t}\cdot\kappa)}{1.615bf_c}\right) \\ & - (1+0.01)^{\mathbf{t}}|M_s| \ . \end{aligned} \tag{6.2}$$

Except for the first phase of structures, where we typically expect the reliability to increase, time-dependent models generally suit the requirements for application of SuSI (monotonicity and independence) naturally. Starting at initial settings (i.e. $t = 0$) with failure probability $3.17 \cdot 10^{-8}$, the failure probability increases quickly with respect to time, having $4.60 \cdot 10^{-7}$, $5.59 \cdot 10^{-6}$, $6.18 \cdot 10^{-5}$, $6.20 \cdot 10^{-4}$ and $4.04 \cdot 10^{-3}$ when considering years $t = 10, 20, 30, 40, 50$. By one simulation run of SuSI, setting the time variable t as dynamic variable $\mathbf{X}_k$, we then get the time-dependent reliability for all desired years. We set the algorithm so that we discover all times for years 1 to 50 (e.g. by assuming a uniform distribution for the time variable, $t \sim U(0, 50)$ in Algorithm 2). An illustration is given in Figure 6.15. The options for formulation of such time-dependent limit states are manifold, allowing to write the limit state function generally as a function of the time t. In particular the option to evaluate several such time-dependent scenarios under different time effects, because single scenario analysis have a low computational demand, allows extensive studies in time-dependent reliability of complex structures.

Note that we could evaluate t in smaller chunks (e.g. days or weeks) at similar costs. Even high demanded accuracies for many single values of t

Figure 6.15: Time-Dependent Reliability Analysis with SuSI. One simulation run yields an estimated failure probability for every year $t = 1, ..., 50$. The algorithm was run with $N = 1000$, $p_0 = 0.3$, and $[p_l, p_u] = [0.2, 0.4]$ and $N^{pred} = 50$, yielding a total of about $E_T = 20,000$ evaluations, including prediction steps, for a complete time-dependent result. Effects by survival of the structure were neglected, deriving unconditional failure probabilities in all cases.



(a) Linear Scale (b) Logarithmic Scale

are easy accessible in SuSI as the bias can be reduced by adding additional grid points or performing a regression on several independent algorithm runs. Efforts still remain similar or only slightly higher with respect to SuS in a static setting.

6.2.3 Robust SuS Evaluations by SuSI

This application is an exceptional example of a SuSI application and should be given high attention, as it does provide a new and eventually very robust approach for general reliability evaluation and is not restricted to the special case of dynamic models. We present how SuSI can be capable of offering a more robust approach on reliability problems, correctly deriving the failure probability in cases where ordinary SuS fails. In Breitung (2018) and Breitung (2019) it was discussed how SuS can be thought of some sort of extrapolation that only yields correct results under specific conditions and how it may be related to a local optimization procedure that will not always find the global optimum, here corresponding to the most important failure regions. The counterexamples where SuS may yield wrong results are based on

the fact that the direction of steepest descent of the samples with respect to their corresponding limit state value does not necessarily lead to exploration of all globally important failure regions. We choose the direction of steepest descent in SuS sampling by successively selecting the subset samples with lowest limit state values in each subset. Reasons for wrong results include existence of several failure regions or the steepest descent at the originating first subsets in the algorithm leading to the wrong directional sampling at first. We show how SuSI can overcome these issues, yielding a robust SuS-like evaluation, most likely independent of the problem.

Examples that Challenge Ergodicity of SuS Sampling (compare Remark 3.2.2)

At first, we select an example from Breitung (2019) for demonstration, considering piecewise linear functions. The procedure is the same for all given examples. The limit state function corresponds to a series system and is given by

$$\min\{g_1(u_1, u_2), g_2(u_1, u_2)\} \ ,$$

with u_1 and u_2 realizations of standard normally distributed random variables U_1, U_2, respectively, and piecewise linear functions

$$g_1(u_1, u_2) = \begin{cases} 4 - u_1 & , u_1 > 3.5 \\ 0.85 - 0.1 \cdot u_1 & , u_1 \leq 3.5 \end{cases}$$

and

$$g_2(u_1, u_2) = \begin{cases} 0.5 - 0.1 \cdot u_2 & , u_2 > 2 \\ 2.3 - u_2 & , u_2 \leq 2 \end{cases} \ .$$

We choose $N = 500$ and $p_0 = 0.1$ for evaluation by SuS. The true failure probability is approximately $p_f = 3.2 \cdot 10^{-5}$, where identifying the wrong failure region as the significant one results in $p_f = 2.9 \cdot 10^{-7}$. The latter thus yields an underestimation of the true failure probability by a factor of 100. SuS will often not find the most important failure region and deliver severely biased results. An illustration is given in Figure 6.16a.

Now SuSI, on the other hand, can be used to circumvent such dangerous shortcomings by manipulating the limit state function. All variables are scaled by a factor of the so-called dynamic variable $\mathbf{X}_k$, where $\mathbf{X}_k$ now plays

the role of an artificial variable that is added to the original problem, resulting in the altered versions

$$g_1^*(u_1, u_2, \mathbf{X}_k) = \begin{cases} 4 - \mathbf{X}_k u_1 & , \mathbf{X}_k u_1 > 3.5 \\ 0.85 - 0.1 \cdot \mathbf{X}_k u_1 & , \mathbf{X}_k u_1 \leq 3.5 \end{cases}$$

and

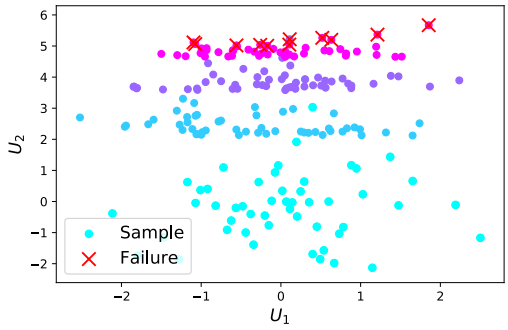
$$g_2^*(u_1, u_2, \mathbf{X}_k) = \begin{cases} 0.5 - 0.1 \cdot \mathbf{X}_k u_2 & , \mathbf{X}_k u_2 > 2 \\ 2.3 - \mathbf{X}_k u_2 & , \mathbf{X}_k u_2 \leq 2 \end{cases}$$

of g_1 and g_2 with corresponding new limit state function

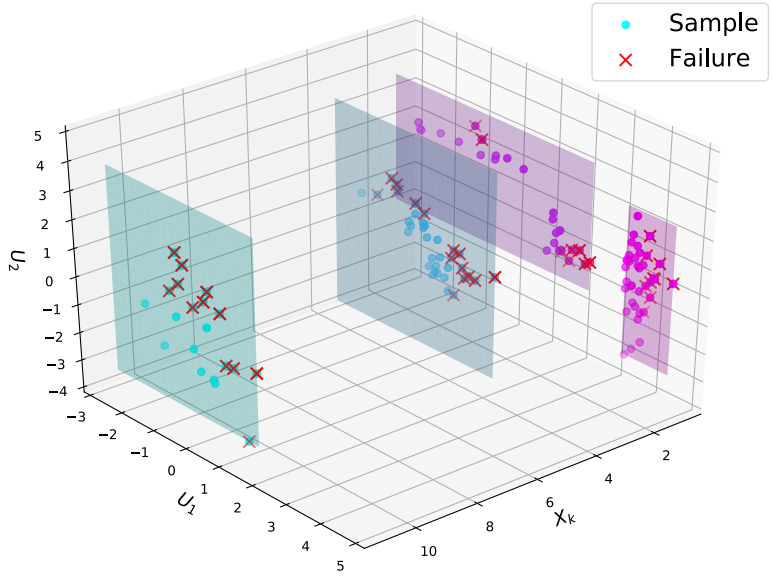
$$\min\{g_1^*(u_1, u_2, \mathbf{X}_k), g_2^*(u_1, u_2, \mathbf{X}_k)\} \ .$$

First, we see that for $\mathbf{X}_k = 1$, this naturally defines the original problem and that for higher $\mathbf{X}_k$, the failure probability will increase. For problems with small failure probabilities, increasing the variance of the model variables will generally result in higher failure probabilities, because then typically the tails of distributions lead to failure. This however, without needing any conditions on the limit state function itself, allows to apply SuSI. The artificially defined $\mathbf{X}_k$ fulfills the SuSI conditions, being independent of all other variables and monotone with respect to the failure probability. Now starting with a high $\mathbf{X}_k$, such as $y_1 = 10 - 15$, the scaling factor is successively decreased until we reach $y_{m^*} \approx 1$ in the last level. Instead of predefining y_1, also nested interval sampling with few samples to find a suitable starting point is appropriate (compare Section 5.2). The procedure is illustrated in Figure 6.16b where also the direct comparison to SuS is given (Figure 6.16). We choose to start by SuS with $y_1 = 11$, $p_0 = 0.1$, $N = 500$ and SuSI with $[p_l, p_u] = [0.2, 0.4]$ and ≈ 150 samples per subset, depending on the exact predicted p_0. These settings result in a total number of evaluations of 3000 for SuS and 3500 for the approach by SuSI, including prediction steps with 50 samples. It is remarkable that SuSI resembles standard Monte Carlo in the first subset, which is also the reason for finding all failure regions, including the dominant region. The result is insensitive to the starting value but should yield a few non-failed samples in the first subset so that the algorithm is applicable in the standard setup. Thus it might be necessary to find a suitable factor by low sample testing in advance, yielding a failure probability less than 100%. However, in our simulations this appeared to be an easy task. According to our implementation (compare Algorithm 2, 'worst' relevant distribution

Figure 6.16: The illustration provides background information on the systematic underestimation of the failure probability by SuS, in contrast to the approach by SuSI. For a clearer view on the subset samples, we only plotted every 10-th sample in SuS as well as every 2nd subset in Figure 6.16a (SuS). For SuSI, we only plot subset levels 1, 2, 5 and 11 in Figure 6.16b and every 5-th sample. The artificial variable $\mathbf{X}_k$ is multiplied to every variable of the original problem, thus drastically altering the failure space in subsets with high $\mathbf{X}_k$, starting with $\mathbf{X}_k = 11$, so that we directly hit the failure region in the first subset by crude Monte Carlo, e.g. transforming $U_2 = 1$ to $U_2 = 11$ in the original limit state function. In a nutshell, we control the geometry of the limit state function to find the desired regions of the sample space, opposed to ordinary SuS, where no controlled manipulation is applied.

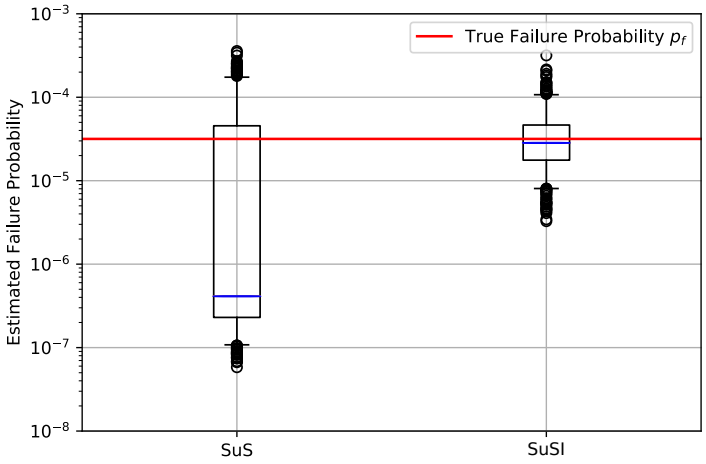


(a) Underestimation of p_f: SuS does not find the most important failure region.



(b) Correct Estimation of p_f: SuSI finds the most important failure region.

Figure 6.17: Illustration of the estimated failure probabilities of SuS and SuSI after 500 independent simulation runs. The whiskers are chosen to contain 90% of all estimates, providing the 95% quantile on the upper side. For simulation, implementation (I4*) was used. In this example, SuS underestimates p_f by a factor of 100 in about 50% of the simulations.



of $\mathbf{X}_k$), sampling according to the needs of this robust approach by SuSI is easily achieved by choosing $\mathbf{X}_k \sim U(1, 11)$ uniformly distributed. More details on the resulting estimates are given in Figure 6.17.

A second example, where SuS dramatically fails to provide correct results is given in Bourinet (2018). The limit state function is given by

$$g(u) = \min_{k \in \{1,2\}} G_k(u)$$

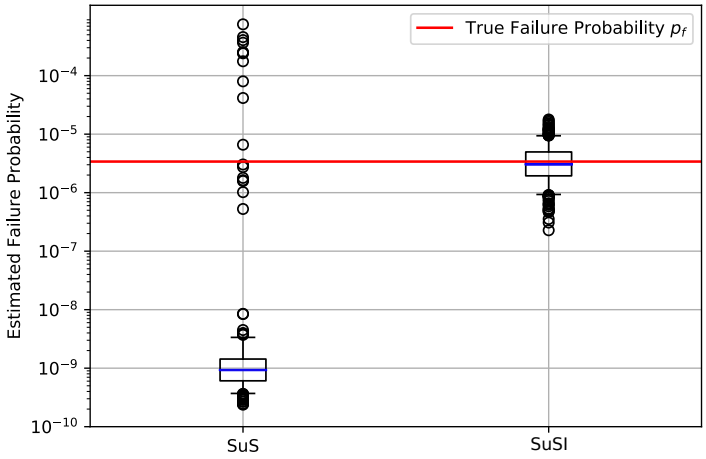
where

$$G_1(u) = (u_1 - 1 \cdot 10^{-6}) + \beta_1 \ ,$$

$$G_2(u) = \beta_1 \left(1 - \left[\frac{1}{2} \left(\frac{u_2 - 1 \cdot 10^{-6}}{\beta_2} + \left| \frac{u_1 - 1 \cdot 10^{-6}}{\beta_2} \right| \right) \right]^{30} \right) \ ,$$

$u = (u_1, u_2)$ two independent standard normal variables and $\beta_1 = 6$, $\beta_2 = 4.5$. The reference failure probability was derived as $p_f = 3.4 \cdot 10^{-6}$ (Bourinet (2018)). In Figure 6.18 we compare the resulting estimates for application of SuS and approach by SuSI for a total number of 8500 and 7000 samples, respectively. Neglecting prediction step samples, e.g. when using repetitive

Figure 6.18: Illustration of the estimated failure probabilities of SuS and SuSI after 500 independent simulation runs. The whiskers are chosen to contain 90% of all estimates, providing the 95% quantile on the upper side. For simulation, implementation (I4) was used. In this example, SuS underestimates p_f by a factor of 1000 in more than 95% of the simulations.



runs for evaluation, would even result in approximately 5000 samples for SuSI. The rRMSE by SuS was found to be given by rRMSE(SuS) = 14.86 in contrast to rRMSE(SuSI) = 0.82, which again indicates the severely biased result when applying SuS. Using SuSI, we in contrast do not have any severe underestimations, always finding the important failure region of the sample space.

The proposed approach should be capable of proving correct results in many cases where SuS leads to severe systematic errors.

Discussion

These are simple examples for illustration and only two dimensions were considered. In such cases, one might detect such issues by SuS. However, in reliability analysis of complex structures with high dimensions, such critical conditions might not be discovered by SuS (Breitung (2018), Breitung (2019)).

The result naturally includes $y_{m^*} = 1$ if it is adequately set up and therefore allows a direct estimation of the original problem, having no bias

in such an evaluation by SuSI. It is also possible to only scale the failure sets with respect to a few variables instead of scaling all of them or initialize with minor or moderate scalings as we can still use preceding SuS as in the original SuSI algorithm. With respect to the optimization analogy (compare Breitung (2018)), repetitive runs with varied scalings can be thought of the counterpart to starting in several initial states to get a better approximation of the globally optimal result.

The presented procedure is, as an application of SuSI, a new one and thus needs further study by testing and analysis. In particular, the possibility to alter limit states in a desired way offers good opportunities for analytical results, as modifications can focus on settings where analytical results can be provided more easily. If at all, we can find similarity to existing methods presented in Bucher (2009) and Qin et al. (2012). Anyhow, there are also many differences between the approaches. We do not need an extrapolation as necessary for these approaches and do explore the domain in a different manner. Still questions such as robustness in high dimensions may arise, since the stated similar method performs bad in such situations. Although such questions need to be answered, such methods should be capable of delivering correct results in most cases, because of its close connection to local and global optimization and a start by crude Monte Carlo for high failure probability estimation. The algorithm starts at N points in a crude Monte Carlo simulation with high enough failure probability and then successively decides which failure events are the most important ones, based on the changing probabilities of the states for decreased $\mathbf{X}_k$. This works very well, as in each subset we just pull away the failure regions from the samples in the speed as it is given by the corresponding real probabilities and since we simulate in the failure regions already given from the start. We assume that wrong results might occur, if failure regions are of the shape of 'small isolated isles', which however are not present if the limit state function is monotone in the variables, and if the pdfs of the random variables of the model have specific distribution functions.

In contrast to SuS, such approaches may not require conditions, or rely only on controllable conditions, on the descent to the failure regions.

Apart from the proposed approach, there are also other ways to get evidence for underestimation of failure probabilities, due to not identifying the correct failure regions. Instead of introducing an artificial variable, one might also successively compute the failure probability under selecting different model variables as $\mathbf{X}_k$, where the results should all coincide. If enough

variables fulfilling the conditions for SuSI are present in the model, then high discrepancies of the conditionally estimated failure probabilities beyond normal variation of the results in the single runs can be considered as an alarming sign, giving evidence for miscalculation.

Conclusion

SuSI appears to provide opportunities for developing new approaches to cope with the problems present in SuS. The presented approach by SuSI seems promising for dealing with problems where SuS yields systematic errors, allowing to carry out more robust reliability analysis.

Remark 6.2.3. *The presented approach for robust evaluation by SuSI can be applied, without any model variable of the original problem formulation fulfilling the SuSI assumptions (monotonicity and independence). In general, the approach should be unbiased, have a similar, although slightly higher, computational demand as SuS (compare Section 6.1, irrelevant* $\mathbf{X}_k$ *case and point estimate, thus approximately doubled computational demands) and provide the opportunity to succeed in reliability analysis where SuS fails.*

6.2.4 Further Considerations and Applications

As we are sampling right at the failure regions in every subset when applying SuSI, we may also attain more information about failure in general. This information could, for example, be utilized to draw conclusions such as evaluating conditionally on deterministic values of $\mathbf{X}_k$ and understanding typical appearances of other stochastic properties of the structure that lead to failure then. The increased knowledge about the failure region may also be used for creating meta models, where in general, meta models, e.g. by utilization of neural networks, could increase efficiency as they may in SuS. Indeed, many enhancements for SuS can as well be applied to SuSI. Another point is the opportunity to also include more conditional variables or trying to relax the conditions on $\mathbf{X}_k$.

There are many more applications to be explored. More examples are also provided in Section 8.2. In particular the opportunity to set $\mathbf{X}_k$ equal to an artificial variable allows for novel creative solutions to existing problems, creating several opportunities for future research.

6.3 Conclusion

In this chapter, we have carried out a simulation study to demonstrate the efficiency of SuSI and illustrated many opportunities for application of the algorithm. The best performance of SuSI was reached for dominant dynamic variables $\mathbf{X}_k$, parameters $[p_l, p_u] = [0.2, 0.4]$ and version b) of the prediction step. Thus, in the use-case of dynamic models, SuSI was found to be very efficient while keeping the remarkable properties of SuS, typically providing computational demands similar as for a static reliability evaluation by SuS. Sometimes, SuSI could even outperform SuS in the static case. The main drawback of SuSI might be its complexity, due to the fact that we are required to perform predictions on unknown probabilities for exploring conditional failure probabilities with respect to $\mathbf{X}_k$ in the prediction step. In some very rare cases, the algorithm might thereby become less stable, more biased or less efficient. Anyhow, if it is well implemented and appropriately applied, we found the challenges to be minor ones compared to the added informative value. Additionally, unstable outcomes could typically be recognized.

A great benefit from the efficient dynamic model evaluation by SuSI is the opportunity to consider novel questions regarding reliability analysis, where SuSI now provides answers in a feasible time. Several applications of the algorithm were demonstrated, showing that it may well satisfy demands in extensive sensitivity and time-dependent reliability analysis. The assumptions, necessary to apply SuSI on dynamic models, are often naturally fulfilled in reliability analysis.

In addition, we showed how SuSI has the potential to compute more robust SuS results or to serve as a supplement for validating results of SuS. The presented robust approach is based on SuSI with respect to an artificial variable and should be generally applicable for most reliability evaluations. In particular at this point, future research should follow, as the presented approach is promising to solve a very important problem in reliability analysis with Subset Simulation.

Chapter 7

Efficient Time-Dependent Reliability Evaluation by a Parameter State Model[1]

This chapter is devoted to a special case of dynamic models, where instead of using SuSI, we preferably use SuS for an efficient analysis. The approach is closely connected to the ideas of SuSI, as it does also rely on the conditional evaluation of reliability with respect to one dynamic variable and the information provided by Subset Simulation. Since SuS can also be understood as a special case of SuSI with less computational demands (compare Example 5.1.4), it is not surprising that there exist dynamic models that should rather be evaluated with SuS directly. We will discuss how such a model must look like and introduce a time-dependent reliability model that suits the requirements. The corresponding model is introduced under the name "parameter state model". It maps all time-dependent information to one stochastic variable, allowing to analyze the original problem, including time-dependent modeling, by analyzing changes in the distribution of this single variable. It is most efficient for, but is not restricted to, Subset Simulation. In particular, this clear-cut structure of our time-dependent model allows for analyzing and visualizing dependencies in detail as well as a novel perspective on reliability estimation, even in complex settings. Additionally, we derive an explicit formula for reliability evaluation, given information of

[1]The contents of this chapter are a slightly altered version of Blandfort et al. (2021). Many parts coincide, but we do not introduce the presented visualization technique here and add references to SuSI.

survival, beyond the typical narrow reliability bounds (e.g. Ditlevsen (1979)). The content of this chapter is the result of a joint work, which has been presented first at Blandfort et al. (2019b) and submitted as an extended version to the Journal of Risk and Uncertainty in Engineering Systems (compare Blandfort et al. (2021)). In contrast to the paper, parameter state model and visualization (see Chapter 8) are split. Additionally, some alterations are made to suit the content of this work better. Although the parameter state model is presented separately from the SuSI algorithm, the information provided by their solutions is identical and can therefore also be presented and understood in the same way, which is demonstrated in Chapter 8.

7.1 Capacity-Demand Factor Model

We start by conversion of the original problem to the capacity-demand model, having separable R and S. This is crucial for the approach by the presented parameter state model. In this section, we state the reliability model as well as our assumptions on S and explain how to efficiently derive a distribution of a possibly complex R.

7.1.1 Structural Reliability

Following Section 2.1, but fixing $b^* = 0$ in this chapter, we have (compare Equation 2.1)

$$p_f = P(g(\mathbf{X}) < 0) = \int_D f_{\mathbf{X}}(\mathbf{x}) 1_{\{g(\mathbf{x})<0\}} d\mathbf{x}$$

for $\mathbf{X} = (X_1, ..., X_d) : \Omega \to D \subseteq \mathbb{R}^d$ being the stochastic properties of the structure, $f_{\mathbf{X}}$ the probability density function (pdf) of X and some limit state function $g : D \to \mathbb{R}$. A function value $g(\mathbf{x})$ smaller than zero indicates failure of the structure for a realization $\mathbf{x}$ of its stochastic properties $\mathbf{X}$. Often, $g(\mathbf{x})$ is rewritten in terms of capacity (or stress resultant) S and resistance R, such that $g(\mathbf{X}) = R(\mathbf{X}) - S(\mathbf{X})$. Assuming stochastic independence between S and R and continuous cumulative distribution function (cdf) F_S of S and pdf f_R of R, yields (Melchers and Beck (2018))

$$p_f = P(S > R) = 1 - \int_{-\infty}^{\infty} f_R(r) F_S(r) dr \overset{(*)}{\approx} 1 - \int_0^{\infty} f_R(r) F_S(r) dr \quad (7.1)$$

and correspondingly its reliability is given by $L = P(S \leq R)$. Note that $(*)$ holds for small $P(R < 0)$, thus for appropriate distributions of R, while always '$\leq$' holds, i.e. $(*)$ does not lead to underestimation of the failure probability.

7.1.2 Properties of S and R

In this section, typical properties of S and R are stated. For formulating the parameter state model and for discussing computational demands, both their distributions need to be derived.

Stress Resultant S. The stress resultant S models the maximum demands over a specific period of time. Suitable distributions are so-called extreme value distributions (De Haan and Ferreira (2007)). The Gumbel distribution is selected here. The following properties are useful:

(G1) If $S_1, ..., S_T$ are Gumbel(μ_i, α) distributed, then $\max\{S_1, ..., S_T\}$ is Gumbel $\left(\alpha \ln\left(\sum_{i=1}^{T} \exp(\mu_i/\alpha)\right), \alpha\right)$-distributed.

(G2) If $S \sim$ Gumbel(μ, α) then $aS + b \sim$ Gumbel$(a\mu + b, a\alpha)$ with $a \in \mathbb{R}^+, b \in \mathbb{R}$.

(G3) If $S \sim$ Gumbel(μ, α) , then its mean value is given by $m_S = \mu + \alpha\gamma$ and its coefficient of variation follows $v_S = \frac{\pi\alpha}{\sqrt{6}m_S}$, with γ the Euler-Mascheroni constant. Vice versa, $\alpha = \frac{m_S\sqrt{6}v_S}{\pi}$ and $\mu = m_S - \alpha\gamma$.

Other extreme value type distributions may be handled similarly, cf. the discussion in Section 7.5.

Resistance R. In contrast to S, there are no restrictions on the distribution of R in the model. This is inspired by applications, where R often contains most of the complexity originally faced in computation of $\{g(\mathbf{X}) < 0\}$, at least computationally. Therefore, the derivation of f_R is the most demanding computational part in (7.1). To derive the pdf f_R, instead of direct computation, first its cdf F_R is assessed. One method to derive F_R is by computing the conditional failure probabilities for a set of deterministic

loads $0 \leq s_1 < s_2 < ... < s_m \in \mathbb{R}^+$, i.e. (note $P(R \leq s_i) = P(R < s_i)$ by continuity)

$$F_R(s_i) = P(R < s_i) = P(R - s_i < 0) = \int_D f_{\mathbf{X}}(\mathbf{x}) 1_{\{R(\mathbf{x}) - s_i < 0\}} d\mathbf{x} \ . \tag{7.2}$$

Note that given a deterministic load s_i, $i = 1, ..., m$, the cdf $F_R(s_i)$ and the conditional failure probability $p_f(s_i) = P(R < S | S = s_i)$ are equivalent, see (7.2). So, standard computational methods for structural reliability with $g_i(\mathbf{x}) := R(\mathbf{x}) - s_i$, $i = 1, ..., n$ are applicable. Interpolation between the points $(s_1, p_f(s_1))$,..., $(s_m, p_f(s_m))$ yields a conditional failure probability function for the whole domain. Sticking to the principle of safety and by monotonicity, $p_f(s_u) = 1$ and $p_f(s_d) = p_f(s_1)$ for all $s_d, s_u \in \mathbb{R}^+$ with $s_d < s_1, s_m < s_u$ are defined. This yields an estimate for the cdf F_R of the resistance R, where estimation sticks to the principle of safety. By taking the derivative of F_R, which works well for interpolation approaches such as monotone splines, also the pdf of R is obtained. Care has to be taken with respect to the interpolation error. Here, using an upper staircase approach instead of smooth interpolation functions will result in an overestimation of the failure probability rather than an underestimation (compare SuSI staircase, e.g. Figure 5.5).

Efficient Computation of F_R for Complex Structures. Calculating several failure probabilities with respect to m deterministic loading values $s_1, ..., s_m$ might be computational demanding, especially when Monte Carlo methods have to be used. To achieve an acceptable interpolation accuracy, often high m are required. However, a method that reduces this effort to a minimum, suiting the requirements perfectly, is Subset Simulation (Au and Beck (2001b), Papaioannou et al. (2015)). Instead of calculating only the failure probability for one given loading distribution, Subset Simulation can calculate the whole distribution of the resistance R by one simulation run. Then, one-dimensional integration yields the failure probability for any loading distribution, as given in (7.1) . This procedure can be extended to more general cases by Subset Simulation Interpolation, see Chapter 5 and Chapter 6. However, here a slightly modified version of Subset Simulation is sufficient and also allows to achieve accurate enough estimates for the cdf of R. The corresponding procedure is presented in Algorithm 3. Note that for evaluation of F_R one can utilize all samples generated in Subset Simulation,

not only the particular subset level results (compare Algorithm 3). A good choice for step 3 (in particular variant a)) is a shape-preserving piecewise cubic Hermite interpolation which guarantees monotonicity of the cdf F_R, see Chapter 5 and Chapter 6.

Note that first and second order reliability methods (FORM, SORM) would need repetitive calculations for every $s_1, ..., s_m$ instead of a single simulation run, so that Subset Simulation is considered a good choice in many applications. Thus, this model is particularly suitable for structures for which Subset Simulation is the preferred method for reliability estimation, such as structures with high dimension or highly non-linear limit states.

7.2 Introduction of the Parameter State Representation

Before adding time-dependence to the model, the parameter state model is introduced in a static setting. The reason for this is that its positive effects are also present in static models, where information about robustness as well as informational visualizations are obtained at a low cost. The parameter state model focuses on keeping the distribution of the resistance in the evaluation constant and represents all dynamic effects by a change of the stress resultant distribution. This way, after pre-calculation of the distribution of R, only a one-dimensional integration, which often has negligible additional costs, is necessary for evaluating (7.1). Indeed, the original complex setting is transformed into a bivariate parameter state, reflecting the distribution parameters of S, so that a new point of view and cheap calculations are the consequence. The resulting state model is defined as follows.

Definition 7.2.1 (State Space)**.** *The state space W consists of all parameter constellations $W = (\mu, \alpha)$ with $\mu \in \mathbb{R}$ the location and $\alpha > 0$ the scale parameter of the Gumbel distribution of $S = S(\mu, \alpha)$.*

Definition 7.2.2 (State Vector)**.** *The state vector $w^* = (\mu^*, \alpha^*) \in W$ describes the current state of the system.*

The object of interest is the corresponding reliability, given a constellation $w^* = (\mu^*, \alpha^*) \in W$ of the system, resulting in reliability (cf. (7.1))

$$L(w^*) = P(S(\mu^*, \alpha^*) \leq R) = \int_0^\infty f_R(r) F_{\mu^*, \alpha^*}(r) dr \ . \tag{7.3}$$

Algorithm 3: Blandfort et al. (2021). Pseudo-Code for Computation of the cdf of R by Subset Simulation

Result: F_R (cdf of R)

Initialization.

$R(\mathbf{x})$ (function for capacity, given a realization $\mathbf{x}$)
$F_{\mathbf{X}}$ (distribution of the stochastic properties of the structure)
$p_{\min}$ (stopping criterion: benchmark for relevant probabilities $p_{\min}$)
N, p_0 (Subset Simulation parameters)
$l = 0$ (current subset level)

1. Initialization of Subset Simulation (Monte Carlo Simulation)

draw N samples $\mathbf{x}_1^0, ..., \mathbf{x}_N^0$ from $\mathbf{X}$
order samples $\mathbf{x}_{(1)}^0, ..., \mathbf{x}_{(N)}^0$ in ascending order
according to their outcomes in capacity $R(\mathbf{x}_1^0), ..., R(\mathbf{x}_N^0)$

2. Subset Simulation

while $p_0^l > p_{\min}$ **do**
  $s_l := R(\mathbf{x}_{(p_0 N)}^l)$
  $F_l^* := \{\mathbf{x} \in D | R(\mathbf{x}) - s_l < 0\}$
  draw N samples $\mathbf{x}_1^l, ..., \mathbf{x}_N^l$ from F_l^* by Markov Chain Monte Carlo
  order samples $\mathbf{x}_{(1)}^l, ..., \mathbf{x}_{(N)}^l$ in ascending order (w.r.t. $R(\mathbf{x})$)
  save all pairs $\left(R(\mathbf{x}_{(p_0 N)}^l), p_0^{l+1}\right)$, $\left(R(\mathbf{x}_{(p_0 N+1)}^l), p_0^l \cdot \frac{p_0 N+1}{N}\right)$,
  ..., $\left(R(\mathbf{x}_{(N)}^l), p_0^l\right)$
  $l = l + 1$
end

3. Estimation of F_R by Interpolation or Smoothing Splines

a) fit a selected model (e.g. spline) to all saved pairs,
considering all levels $i = 0, ..., l$
b) fit a selected model to a subset of the evaluated pairs
The resulting function is the estimated cdf $F_R(r)$ of R

In this model, calculation of the pdf f_R allows evaluation of reliability by (7.3) for arbitrary choices of parameters of the loading variable. Note that different types of distributions of the loading variable can be applied as well, although then the comparison between different loading distributions in visualization becomes harder (for details see Chapter 8, especially Figure 8.3).

7.3 Time-Dependent Reliability Analysis

In this section, time-dependent effects are added to the model, switching to a time-dependent reliability formulation.

Initial, Current and Future Time. First, $t = 0$ refers to the time when the structure was built and initially put into operation. All other relevant times then are positive real numbers $t > 0$. At each time $t > 0$, a stochastic real-valued resistance R_t and stress resultant S_t are given. Those induce the real-valued stochastic processes $(R_t)_{t\geq 0}$ and $(S_t)_{t\geq 0}$. Further, let t_c define the current and t_n the final relevant time. The task in reliability estimation is then to estimate the failure probability of some future time interval $(t_k, t_l]$ for times $t_c \leq t_k < t_l \leq t_n$ (e.g. $t_l - t_k = 50$ years, $t_k = t_c$).

Stopping Time for Failure of the Structure. Additionally, the stopping time $\tau_F := \min\{t \in \mathbb{R}^+ \cup \{\infty\} : R_t < S_t\}$ is introduced, defining the (first) time when failure of the structure occurs, where ∞ means that no failure ever occurs. Also set $R_t = 0$ for all $t \geq \tau_F$, meaning that after failure occurred the structure has broken down and can no longer handle any demands, remaining in a failed state. If one wants to estimate the reliability of a structure, it has not failed so far. Thus it is convenient to assume $\tau_F > t_c$.

7.3.1 Time Effects

Discretization. The discretized approach is used for time-dependent reliability. The observation period $(0, t_n]$ is split into $n > 1$ smaller equidistant disjoint time periods. In more detail, a discretization induced by the set of time instances $\mathbb{T} = \{t_0, t_1, t_2, ...t_n\}$ with $0 = t_0 < t_1 < ... < t_n$ and $t_j - t_{j-1} = c$ for all $j = 1, ..., n$ and some $c \in \mathbb{R}$ is chosen. This discretization splits the observation period naturally into n disjoint time intervals $(0, t_n] = I_1 \cup I_2 \cup ... \cup I_n$ for $I_j := (t_{j-1}, t_j]$, $j = 1, ..., n$. The interval length

is c and the set of intervals is referred to as $\mathbb{I} := \{I_1, ..., I_n\}$. To stick to the principle of safety with respect to time continuity, the reliability of time period $I \in \mathbb{I}$ is determined by

$$L(I) = P(\max\{S_t : t \in I\} \leq \min\{R_t : t \in I\}) \ . \tag{7.4}$$

This way, the failure probability is not underestimated. This induces how to choose the time effects for non-stationary maximum loading distribution and changes in resistance of the structure, if given the corresponding time-continuous processes.

Deterioration and Post-Curing. Both effects have an impact on the resistance of the structure, which is defined by a time-dependent function $\theta_{cr} : \mathbb{T} \to \mathbb{R}^+, \theta_{cr}(0) = 1$. It is assumed that this function is decomposable into its components, i.e. $\theta_{cr}(t) = \theta_c(t)\theta_r(t)$ for deterioration effect function θ_r decreasing in t and post-curing function θ_c increasing in t, so that the resistance is given by $R_t = \theta_c(t)\theta_r(t)R_0$, with R_0 the initial resistance. In interval representation, these functions must yield the minimum resistance of the whole interval. Therefore the corresponding interval version of the time effects is defined as $\theta^I_{cr} : \mathbb{I} \to \mathbb{R}^+, \theta^I_{cr}(I) := \min\{\theta_{cr}(t) : t \in I\}$. θ^I_c and θ^I_r are derived from the time continuous representatives by taking the same minimal point t as in $\theta^I_{cr}(I)$ such that $R_I = \theta^I_{cr}(I)R_0 = \theta^I_c(I)\theta^I_r(I)R_0$ still holds.

Non-Stationary Loadings. In contrast to the resistance and effects on it, the maximum loading distribution and its change in distribution naturally refer to an interval. The mean value of the maximum loadings changes according to the function $\theta^I_s : \mathbb{I} \to \mathbb{R}, \theta^I_s(I_1) = 1$ so that $S_I \sim \theta^I_s(I)S_{I_1}$ for $I \in \mathbb{I}$.

Survival Effect. If the structure has already survived for a specific time, the resistance of the structure should be updated based on this information. Resistance at time t, given information of survival up to time s, equivalently given by $\tau_F > s$, is then written shortly as $R_{t|\tau_F>s}$. The interval representation is written as $R_{I_l|\tau_F>t_k}$, $1 < k \leq l < n$ referring to the minimum reliability in interval I_l given survival in $I_1, ..., I_k$. Typically, survival up to the last time period $R_{I_l|\tau_F>t_{l-1}}$ is needed for successive analysis of reliability over time. Note that by assumption of independence between maximum

loadings of disjoint intervals, there is no gain in information about the stress resultant, hence $S_{I_j|\tau_F>t_{j-1}} = S_{I_j}$. The survival effect needs the most computational effort and has a special role in the model. Also note that for conservative estimation, in contrast to computation of reliability, one should preferably switch maximum and minimum in comparison to (7.4) for derivation of $R_{I_l|\tau_F>t_k}$. Thus the corresponding time effects for each of the past intervals are generally different. This is the case, since otherwise the effect given by survival in the past could be overestimated, lowering future reliability estimates too much. However, in the following this effect is neglected since it seems to be not that severe here. Its inclusion is straightforward and should be considered if very accurate results including the survival effect are necessary or time intervals are long.

Resulting Effect. In conclusion, the model includes effects by discretization, deterioration, post-curing, non-stationary loadings and survival of the structure. Except for survival of the structure, all effects are assumed linear. Assuming R and S take their worst interval values at the same time, this leads to an overestimation of failure probability. However, shortening the interval length reduces this approximation error. Note that fatigue effects do not suit the model requirements due to the decoupled $R - S$ model. In the following each of those effects is further explored and it is shown how to integrate them into the parameter state model.

7.3.2 Integration of the Time Effects in the Parameter State Model

This section deals with the representation of the above time effects in the parameter state model. To do so, one has to transform time-dependent reliability formulas for single periods with non-identical distributions of resistance and stress resultant into formulas with identical pdfs of R and fixed distribution type of the stress resultant. In brief, the aim is to rewrite the reliability

$$L_I = P(S_I \leq R_I) = \int_0^\infty f_{R_I}(r) F_{S_I}(r) dr$$

for time intervals $I \in \mathbb{I}$ in terms of a state $w(I) = (\mu(I), \alpha(I)) \in W$ of the parameter state model. Again sticking to the principle of safety, the 'worst' state of the interval I is taken as a representative. For initial resistance R_0

and some loading distribution $F_{S_I} = F_{\mu(I),\alpha(I)}$, one has

$$L_I = P(S_I \leq R_I) = \int_0^\infty f_{R_0}(r) F_{\mu(I),\alpha(I)}(r) dr \ . \tag{7.5}$$

The properties of extreme value distributions fit perfectly here, cf. Section 7.5.

Discretization. The max stable property (G1) guarantees that splitting or merging of time periods is consistent, also yielding an explicit formula for meaningful resulting stress resultant distributions and the opportunity to easily reduce interval length.

Non-Stationary Loadings. Those are themselves just a direct change of the loading distribution $F_{\mu,\alpha}$. In particular $S_I \sim \theta_s^I(I) S_{I_1}$ allows to apply (G2) for an explicit formula to get the new state parameters.

Deterioration and Post-Curing. Considering both effects, deterioration and post-curing, results in

$$L_I = P(S_I \leq R_I) = P(S_I \leq \theta_c^I(I)\theta_r^I(I) R_0) = P\left(\frac{S_I}{\theta_c^I(I)\theta_r^I(I)} \leq R_0 \right) \ .$$

So, the location scale property (G2) yields equivalence of the effects on resistance and a change in the parameters $\mu(I)$ and $\alpha(I)$ of the loading distribution (7.5).

Survival Effect. This effect is the most difficult to integrate into the model as it has to account for the past. Thus, Section 7.4 is devoted to the derivation of its impact on reliability only. The resulting reliability in period $I_j \in \mathbb{I}$ given survival up to the previous period I_{j-1}, or equivalently time t_{j-1}, is then given as

$$L_{I_j|\tau_F > t_{j-1}} = \int_0^\infty \frac{\prod_{i=1}^{j} F_{S_{I_i}}\left(s \cdot \theta_{cr}^I(I_i)\right)}{\int_0^\infty \left(\prod_{i=1}^{j-1} F_{S_{I_i}}\left(v \cdot \theta_{cr}^I(I_i)\right) \right) f_{R_0}(v) dv} f_{R_0}(s) ds \ . \tag{7.6}$$

A proof is given in Section 7.4. The survival effect does not allow to give explicit formulas for the state of the system anymore, but still does for the reliability. For illustrations, it is thus often not included in the model first, allowing for a more straightforward analysis and presentation.

Resulting Effect. Concluding, the state vector for time period $I \in \mathbb{I}$, without survival effect, is given by

$$w(I) = \left(\mu(1) \frac{\theta_s^I(I)}{\theta_c^I(I)\theta_r^I(I)}, \alpha(1) \frac{\theta_s^I(I)}{\theta_c^I(I)\theta_r^I(I)} \right) .$$

The initial state $\mu(1), \alpha(1)$ corresponds to the maximum loading distribution $F_{S_{I_1}}$ of the first period I_1. Note that both parameters are multiplied by the same factor, resulting in a constant coefficient of variation. A main advantage of this approach is its flexibility regardless of how complex the single components of the factor become or if they are assumed to be stochastic, the model remains simple and computational efforts are only slightly increased. In the following, this type of model is referred to as 'factor model'. Although this model is a bit restrictive and an extension to general linear effects on capacity and arbitrary effects on demand is possible, it simplifies analysis significantly and allows for an even better understanding.

Under the survival effect, the computational time is still not significantly effected for complex structures because still only one dimensional integration is necessary for reliability estimation as given in (7.6). On the other hand, with survival effect it is not a priori clear how to find a representative state. Deriving the best fitting state representation now requires to take a numerical approach as in many cases no explicit solution exists, as

$$\frac{\prod_{i=1}^{j} F_{S_{I_i}} \left(s \cdot \theta_{cr}^I(i) \right)}{\int_0^\infty \left(\prod_{i=1}^{j-1} F_{S_{I_i}} \left(v \cdot \theta_{cr}^I(i) \right) \right) f_{R_0}(v) dv}$$

in general does not belong to the family of extreme value distributions anymore. One way to circumvent this problem is to search for parameter combinations $(\mu^*, \alpha^*) \in \mathbb{R} \times \mathbb{R}^+$ of the loading distribution that yield the same as the beforehand computed reliability $L_{I_j|\tau_F > t_{j-1}}$. However, as there exist infinitely many solutions, which yield equality in reliability, another reasonable criterion has to be specified. Our proposal is to minimize the Euclidean distance of the state (μ^*, α^*) to the state under exclusion of the survival effect (μ^b, α^b) which is available by following the standard procedure. Allowing an error of $\epsilon > 0$ in reliability, this results in

$$\begin{array}{ll} \text{minimize} & (\mu^* - \mu^b)^2 + (\alpha^* - \alpha^b)^2 \ , \\ \text{subject to} & \left| \int_0^\infty f_{R_0}(r) F_{\mu^*, \alpha^*}(r) dr - L_{I_j|\tau_F > t_{j-1}} \right| \leq \epsilon \ . \end{array}$$

Again, finding the solution does not have a significant computational demand if complex structures are analyzed.

Note that monitoring and inspection can also be included if the corresponding information has a linear effect on R. If given a specific guess for the cross sectional area of the steel, not knowing how it will effect R relatively at all, more sophisticated approaches are necessary.

7.4 Reliability Under the Survival Effect: Derivation of the Explicit Formula

This section derives the reliability under consideration of the survival effect, as given in (7.6). For ease of notation, j is written instead of $I_j = (t_{j-1}, t_j)$ for reference to time intervals and $R_{j|j-1}$ instead of $R_{I_j|\tau_F > t_{j-1}}$ for known survival up to the previous time period. The discretized model is kept, so all references with respect to time, except for the initial resistance R_0 and its pdf f_{R_0}, correspond to time intervals. Note, as described in Section 7.3, for perfect alignment one would need other time effect functions for the intervals. To derive the explicit formula for reliability, in a first step, it is shown that in the time-dependent factor model, the updated pdf of the resistance for period $j \geq 2$ with information of survival in period $j-1$ is given by

$$f_{R_{j|j-1}}(r) = \frac{\left(\prod_{i=1}^{j-1} F_{S_i}\left(\frac{r}{\prod_{k=i+1}^{j}\Delta_k}\right)\right) f_{R_0}\left(\frac{r}{\prod_{k=1}^{j}\Delta_k}\right)}{\int_0^\infty \left(\prod_{i=1}^{j-1} F_{S_i}\left(\frac{u}{\prod_{k=i+1}^{j}\Delta_k}\right)\right) f_{R_0}\left(\frac{u}{\prod_{k=1}^{j}\Delta_k}\right) du} \,. \tag{7.7}$$

For ease of notation also define $\Delta_j := \frac{\theta^I_{cr}(j)}{\theta^I_{cr}(j-1)}$, the relative change of resistance from time interval $j-1$ to j. In the following, a proof of (7.7) is given by induction on t for arbitrary $r \in \mathbb{R}$.

Base Case $j = 2$: By definition of the pdf, non-formally, one has $f_{R_{1|1}}(r) = \frac{1}{d\tilde{r}} P(r < R_1 < r + d\tilde{r} | S_1 \leq R_1)$ for an infinitely small $d\tilde{r} \in \mathbb{R}^+$. Then Bayes

Formula yields

$$\frac{1}{d\tilde{r}}P(r < R_1 < r + d\tilde{r}|S_1 \leq R_1) \\ = \frac{P(S_1 \leq R_1|r < R_1 < r + d\tilde{r})P(r < R_1 < r + d\tilde{r})}{d\tilde{r}\int_0^\infty P(S_1 \leq R_1|R_1 = s)f_{R_{1|0}}(s)ds} \ . \quad (7.8)$$

$P(S_1 \leq R_1|r < R_1 < r + d\tilde{r}) = F_{S_1}(r)$ and $P(r < R_1 < r + d\tilde{r}) = f_{R_1}(r)d\tilde{r}$, therefore yields

$$f_{R_{1|1}}(r) = \frac{F_{S_1}(r)f_{R_{1|0}}(r)}{\int_0^\infty F_{S_1}(s)f_{R_{1|0}}(s)ds} \ .$$

Next, the cdf of the transformed random variable $R_{2|1} = \Delta_2 R_{1|1}$ is given by

$$F_{R_{2|1}}(r) = P(R_{2|1} \leq r) = P(\Delta_2 R_{1|1} \leq r) \\ = P\left(R_{1|1} \leq \frac{r}{\Delta_2}\right) = F_{R_{1|1}}\left(\frac{r}{\Delta_2}\right) \ .$$

Thus for the pdf

$$f_{R_{2|1}}(r) = \frac{dF_{R_{1|1}}(\frac{r}{\Delta_2})}{dr} = \frac{1}{\Delta_2}f_{R_{1|1}}\left(\frac{r}{\Delta_2}\right) \\ \overset{(7.8)}{=} \frac{1}{\Delta_2}\frac{F_{S_1}\left(\frac{r}{\Delta_2}\right)f_{R_{1|0}}\left(\frac{r}{\Delta_2}\right)}{\int_0^\infty F_{S_1}(s)f_{R_{1|0}}(s)ds} \ .$$

Then $f_{R_{1|0}} = \frac{1}{\Delta_1}f_{R_0}(\frac{r}{\Delta_1})$ and substitution of $u = s \cdot \Delta_2$ gives

$$f_{R_{2|1}}(r) = \frac{1}{\Delta_2}\frac{F_{S_1}\left(\frac{r}{\Delta_2}\right)f_{R_{1|0}}\left(\frac{r}{\Delta_1\Delta_2}\right)}{\int_0^\infty F_{S_1}(s)f_{R_{1|0}}\left(\frac{s}{\Delta_1}\right)ds} \\ = \frac{F_{S_1}\left(\frac{r}{\Delta_2}\right)f_{R_0}\left(\frac{r}{\Delta_1\Delta_2}\right)}{\int_0^\infty F_{S_1}\left(\frac{u}{\Delta_2}\right)f_{R_0}\left(\frac{u}{\Delta_1\Delta_2}\right)du} \ .$$

Thus the claim holds true for $j = 2$.

Induction step $j \to j+1$**:** As above, by Bayes Formula

$$\begin{aligned} f_{R_{j-1|j-1}}(r) &= \frac{P(S_{j-1} \leq R_{j-1} | R_{j-1} = r) f_{R_{j-1|j-2}}(r)}{\int_0^\infty P(S_{j-1} \leq R_{j-1} | R_{j-1} = s) f_{R_{j-1|j-2}}(s) ds} \\ &= \frac{F_{S_{j-1}}(r) f_{R_{j-1|j-2}}(r)}{\int_0^\infty F_{S_{j-1}}(s) f_{R_{j-1|j-2}}(s) ds} \ . \end{aligned} \tag{7.9}$$

Next using the transformation rule, as in the base case, again, results in

$$f_{R_{j|j-1}}(r) = \frac{1}{\Delta_j} f_{R_{j-1|j-1}}\left(\frac{r}{\Delta_j}\right) \overset{(7.9)}{=} \frac{1}{\Delta_j} \frac{F_{S_{j-1}}\left(\frac{r}{\Delta_j}\right) f_{R_{j-1|j-2}}\left(\frac{r}{\Delta_j}\right)}{\int_0^\infty F_{S_{j-1}}(s) f_{R_{j-1|j-2}}(s) ds} \ .$$

Substitution in the integrand by $u := s\Delta_j$ yields

$$f_{R_{j|j-1}}(r) = \frac{1}{\Delta_j} \frac{F_{S_{j-1}}\left(\frac{r}{\Delta_j}\right) f_{R_{j-1|j-2}}\left(\frac{r}{\Delta_j}\right)}{\int_0^\infty F_{S_{j-1}}\left(\frac{u}{\Delta_j}\right) f_{R_{j-1|j-2}}\left(\frac{u}{\Delta_j}\right) du} \ .$$

Then, by the induction hypothesis, $f_{R_{j|j-1}}(r)$ is equal to

$$\frac{F_{S_{j-1}}\left(\frac{r}{\Delta_j}\right)\left(\prod_{i=1}^{j-2} F_{S_i}\left(\frac{r}{\prod_{k=i+1}^{j-1}\Delta_k \Delta_j}\right)\right) f_{R_0}\left(\frac{r}{\prod_{k=1}^{j-1}\Delta_k \Delta_j}\right)}{\int_0^\infty F_{S_{j-1}}\left(\frac{u}{\Delta_j}\right)\left(\prod_{i=1}^{j-2} F_{S_i}\left(\frac{u}{\prod_{k=i+1}^{j-1}\Delta_k \Delta_j}\right)\right) f_{R_0}\left(\frac{u}{\prod_{k=1}^{j-1}\Delta_k \Delta_j}\right) du}$$

which yields (7.7). So (7.7) was shown for $j \geq 2$.

Now it remains to derive the corresponding interval reliability by the pdf $f_{R_{j|j-1}}$. The reliability in time period $j \geq 2$, given survival in the previous time periods $1, 2, ..., j-1$ is given by

$$L_{j|j-1} = \int_0^\infty \frac{\prod_{i=1}^{j} F_{S_i}(s \cdot \theta^I_{cr}(i))}{\int_0^\infty \left(\prod_{i=1}^{j-1} F_{S_i}(v \cdot \theta^I_{cr}(i))\right) f_{R_0}(v) dv} f_{R_0}(s) ds \ . \tag{7.10}$$

This can be shown in a straightforward way, because

$$\begin{aligned} L_{j|j-1} &= \int_0^\infty f_{R_{j|j-1}}(r) F_{S_j}(r) dr \\ &\overset{(7.7)}{=} \int_0^\infty \frac{\left(\prod_{i=1}^{j-1} F_{S_i}\left(\frac{r}{\prod_{k=i+1}^{j}\Delta_k}\right)\right) f_{R_0}\left(\frac{r}{\prod_{k=1}^{j}\Delta_k}\right)}{\int_0^\infty \left(\prod_{i=1}^{j-1} F_{S_i}\left(\frac{u}{\prod_{k=i+1}^{j}\Delta_k}\right)\right) f_{R_0}\left(\frac{u}{\prod_{k=1}^{j}\Delta_k}\right) du} F_{S_j}(r) dr \end{aligned}$$

with $\prod_{k=1}^{j} \Delta_k = \theta_{cr}^I(j)$. Substitution of $s = \frac{r}{\theta_{cr}^I(j)}$ and $v = \frac{u}{\theta_{cr}^I(j)}$ then yields

$$L_{j|j-1} = \int_0^\infty \frac{\left(\prod_{i=1}^{j-1} F_{S_i}(s \cdot \theta_{cr}^I(i)\right) f_{R_0}(s)}{\int_0^\infty \left(\prod_{i=1}^{j-1} F_{S_i}(v \cdot \theta_{cr}^I(i)\right) f_{R_0}(v) dv} F_{S_j}(s \cdot \theta_{cr}^I(j)) ds$$

which provides (7.10) by rearranging terms.

7.5 Applicability of the Model

Model Assumptions and Requirements. The presented approach is limited to specific structural models, the main constraints are:

- In the static case, the reliability model is a capacity-demand (R-S) model with independent R and S, in which demands are given as a single stochastic or deterministic variable.
- If time-dependent effects are included and the demand variable is stochastic, its distribution has to belong to the location-scale family (e.g. Gaussian, Cauchy, Uniform, Gumbel).
- Time-effects on R need to be linear and scalar for the bar plot representation. The reliability formula with consideration of survival was derived in Section 7.4 for scalar effects only.
- If transformations between different reference periods (say, from fifty years to one year) are desirable, S should have an extreme value distribution (Fréchet, Weibull, Gumbel).

Examples. Examples are given in Section 8.1 under inclusion of the visualization technique of Chapter 8. Among other engineering structures, it is applied to a network problem which allows for an arbitrary dimension in general. Although the model is restrictive, there exist many cases of complex structures where its requirements are fulfilled. For example the complex reliability assessment of the ultimate compressive strength of stiffened plate elements in Gaspar et al. (2014) can easily be brought in line with the model assumptions. For that example, it is sufficient to transform the sum of two bending moments (Normal and Gumbel distributed, respectively) into a single Gumbel distributed stress resultant. Then the model has the desired

form. The resistance of the structural model needs to be evaluated by finite element analysis in this example, so that its complexity and computational demands are exceptionally high. Albeit this high complexity, after evaluation of resistance by Algorithm 3, one can still consider effects such as a linear decrease in resistance or a change of the stress resultant, in general or over time, without relevant additional computational demands.

Model Extensions. Furthermore, the model requirements for S and R can be switched in the model, allowing to analyze more types of complex structures such as oscillating systems (e.g. two degree of freedom primary/secondary damped oscillator in Bourinet et al. (2011)), or a transmission line studied in Kouassi et al. (2016), where complexity of the model is concentrated in S instead of R. This extends the applicability of the proposed model structure. As an example, we consider the transmission line (Kouassi et al. (2016)) in Chapter 8.

Depending on the specific setting, a model fulfilling the requirements may also serve as an approximation to more general structural models. Then, accuracy in the approximated model is traded in for efficiency in the approximating model, the latter allowing a more extended analysis with respect to robustness and visualization. To find a good balance here, depends on the applications under consideration.

7.6 Conclusion

A novel parameter state model for extensive time-dependent reliability analysis was presented. Using mathematical properties of the classical capacity-demand model and of extreme value distributions, it was shown how to compute time-dependent reliability over a whole reference time interval at similar cost as a static reliability evaluation with Subset Simulation. The proposed procedure is independent of the structure of R, given that the model requirements are fulfilled. The computational benefit lies mainly in the fact that after evaluation of R, i.e. after estimation of the cdf F_R, in the parameter state model one can analyze all proposed time-dependent aspects at negligible costs.

Considerable computational benefits are achieved, if the complexity of the capacity is exceptionally high such as in the given network example, presented in Chapter 8, with a high number of nodes or if demanding finite

element analysis is necessary. After analyzing R one can again proceed as above to obtain further properties at low costs.

The drawback of the proposed parameter state model is the restriction to its specific model requirements. However, many relevant examples are covered and model extensions are possible as discussed in Section 7.5.

Furthermore, an explicit formula for efficient calculation of time-dependent reliability under its survival history was derived, allowing for better approximations than by often used classical narrow bounds for reliability.

Concluding, the main benefits of the proposed approach are threefold:

- The opportunity to conduct intensive case studies for complex structures at a feasible cost, potentially at the order of an ordinary static reliability evaluation, allows for finding new ways of altering design and also for identifying dangers which may remain unseen otherwise.

- The model allows for efficient calculation of reliability given past survival of a structure by an explicit formula.

- In Chapter 8, we show how informative visualizations by heat maps and bar charts yield a holistic and robust view at time-dependent reliability models when using the parameter state model.

Chapter 8

Informative Visualization[1]

In reliability analysis, results of dynamic models, and in particular time-dependent reliability analysis, are mostly displayed as reliability over time and plots thereof. Although this representation yields important information and there are extensions such as Bayesian networks (Straub (2009), Straub and Der Kiureghian (2010), Luque and Straub (2016), Zwirglmaier and Straub (2016)), or illustrations based on outcrossing (e.g. Melchers and Beck (2018)), it seems that generally new forms of comprehensibility are often not taken into consideration much. However, it is an important task to present results in a understandable and informative way, especially in the case of complex structures. Thus, this chapter is dedicated to the presentation of informative visualization techniques, which yield a holistic and robust view at reliability analysis.

Three engineering structures are chosen for demonstration of such visualizations when using the parameter state model as well as when applying SuSI, respectively. We aim at showing how the developed approaches allow for efficient extensive reliability studies of complex structures and how one can create informative visualizations, perfectly suitable for the approaches presented in this dissertation. Following Blandfort et al. (2019b) and Blandfort et al. (2021), we use a new visualization technique that allows for a clear view on model uncertainties and dynamics, applicable for both SuS and SuSI. On the one hand, SuSI is rather generally applicable and examines single dynamic variables in great detail, the parameter state model, on the other hand, requires specific conditions but additionally addresses the

[1] The content of this chapter is partially an extract of Blandfort et al. (2021), containing but also extending the visualization part of it and applying it on SuSI additionally.

modeling of time-dependent factors and is slightly more efficient. This suits the identification of SuS as a special case of SuSI, as shown in Example 5.1.4. Note that not all considered structures are complex enough to require using SuS for reliability evaluation. Nevertheless, due to the very high number of necessary evaluations for drawing the presented heat maps, it might be useful to use the given approaches. This work focuses on efficient computation of dynamic models, but also on extension of perspectives and understanding of reliability evaluation. Therefore, this chapter provides a visualization concept, specifically designed for our models but also suitable for reliability analysis in general, that allows to show lots of information and to include dynamic models of many forms. Also, it naturally includes analysis of uncertainties and offers information such as distance of the current distribution parameters of the dynamic variable to a given reliability benchmark.

Visualization: Based on Conditional Failure Probabilities

Our novel algorithm SuSI (Chapter 5-6) and the parameter state model (Chapter 7) both provide extended information on reliability evaluation for dynamic models. Both are based on a conditional evaluation with respect to a dynamic variable $\mathbf{X}_k$, that can be chosen almost fully arbitrary in SuSI and requires to be the stress resultant variable S (or capacity/resistance R respectively) of an independent capacity-demand model for application of the parameter state model. Now, for this dynamic variable $\mathbf{X}_k$, many distributions may be considered without a relevant additional computational demand. In general, one could make a list of results for different distributions to provide the manifold result. However, such a result is difficult to interpret and one tempts to look at less constellations and misses analysis of some interesting ones, if specific distributions have to be selected individually. For that reason, we use a specialized visualization that suits the novel results, showing lots of information in an easy understandable way. Furthermore, it is desirable to offer a new perspective on the reliability problem according to the extended information provided by the results. We start by consideration of the parameter state model to introduce the procedure and then demonstrate how it also suits SuSI perfectly.

8.1 Parameter State Model

To demonstrate the parameter state model and also the corresponding visualization of its results, we consider three engineering structures. The first is a simple 3-span reinforced concrete slab which allows for an easy understanding of the procedure. Then, a network is analyzed, showing the potential for analysis of structures that potentially take any size or complexity. Lastly, a transmission line is considered, demonstrating that it is admissible to switch the roles of capacity and demand as stated in Section 7.5.

8.1.1 3-Span Reinforced Concrete Slab (cont.)[2]

Static Model

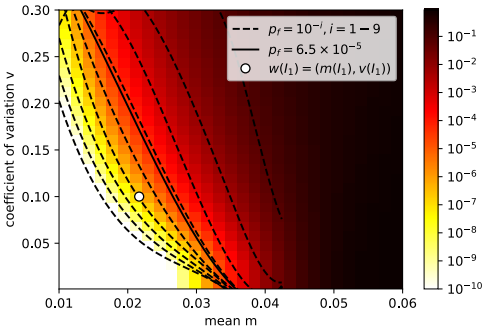
For demonstration of the visualization technique and the parameter state model, we continue analyzing the reliability of the static 3-span reinforced concrete slab introduced in Section 6.2.2. This simple example provides an easy understanding of the procedure. An illustration was given in Figure 6.14a where Table 6.14c contains the corresponding stochastic properties. With Subset Simulation, the cdf F_R of

$$R = U_M(f_y A_s)\left((h - d_1) - \frac{f_y A_s}{1.615 b f_c}\right)$$

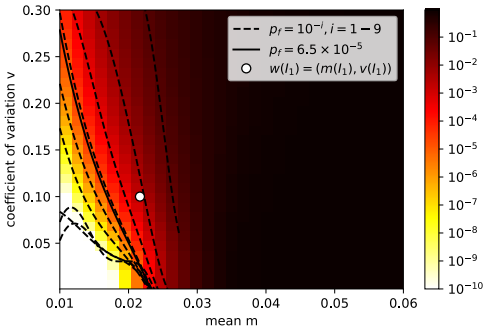
is computed by one simulation run (as described in Algorithm 3). One may compute failure probabilities for arbitrary distributions of the resulting moment force M_s then. Accordingly, the state space W of $|M_s|$ is given by the parameters μ, α of the Gumbel distributed random variable $|M_s|$ with static state $(\mu^s, \alpha^s) = (0.02069, 0.00169)$, which is equivalent to $(m^s, v^s) = (0.02166, 0.1)$ by (G3), for 50 years of reference time. Using (7.1), we span a 30×30 grid for the states, covering $(m_S, v_S) \in [0.001, 0.06] \times [0.001, 0.30]$ with a total of 900 different parameter constellations and assign a specific color to each of the resulting estimated failure probabilities.

The result is shown in Figure 8.1a. Note that the repetitive evaluations of the one-dimensional integration according to (7.1) have a negligible computational demand compared to typical structural reliability evaluations of complex structures. Thus, the effort is similar as for a single reliability evaluation, providing the potential for drastically reducing computational

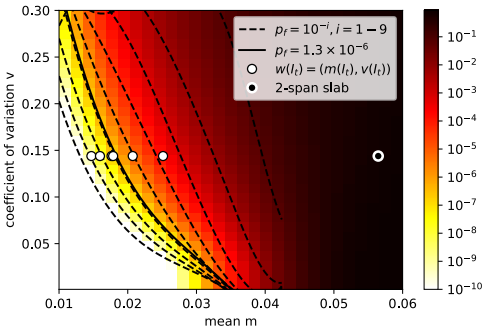
[2]Based on Blandfort et al. (2021).



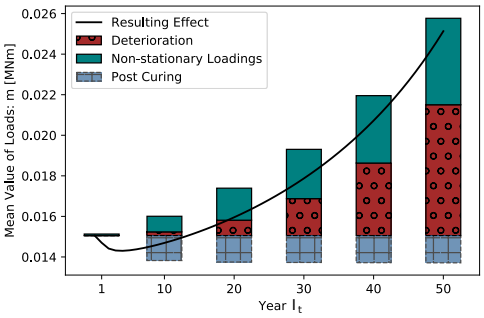
(a) $A_s = 0.000524 \frac{m^2}{m}$



(b) $A_s = 0.000335 \frac{m^2}{m}$



(c) Time-Dependent Model



(d) Time-Dependent Model: Bar Chart

Figure 8.1: Blandfort et al. (2021). Analysis of the 3-span slab example (compare Figure 6.14) in the parameter state model, using heat maps and a bar chart. A comparison between Figure 8.1a and Figure 8.1b demonstrates the effect of a changed steel cross sectional on the appearance of the heat map. Analysis of the time-dependent model is realized in Figure 8.1c and Figure 8.1d, where the former allows conclusions with respect to reliability and the bar chart decomposes the effects which lead to a change of the time-dependent parameter state in it, allowing to identify causes and relations of change in reliability over time.

demands in such extensive analyses. The solid curve in Figure 8.1a marks the target reliability for fifty years according to Eurocode DIN EN 1990: 2010-12 (2010), while the dotted curves are additional supportive lines for better readability. The curves are derived by a two-dimensional spline interpolation. Even without including time effects, one already benefits by a clear view on the behavior of reliability with respect to uncertainties or dynamics in the loading variable, yielding a very detailed sensitivity analysis that goes beyond classical local predictions. Additionally, a heat map for the case $A_s = 0.000335\frac{m^2}{m}$ is given in Figure 8.1b, to demonstrate the impact of a fundamental change of resistance R by a single altered stochastic variable.

Time-Dependent Model

The evolution of the state $w(t)$ with respect to time is considered. Now it is assumed that resistance and loading distributions evolve over time. At first, the original problem with fifty years reference period is discretized. By (G1), a consistent transformation of the 50 year maximum distribution of the loadings with parameters $(\mu_{50}, \alpha_{50}) = (0.02069, 0.00169)$ or correspondingly $(m_{50}, v_{50}) = (0.02166, 0.1)$ to a yearly maximum loading distribution, if no time effects are present, is derived by $\mu_1 = \alpha_{50} \log\left(\frac{1}{50}\exp\left(\frac{\mu_{50}}{\alpha_{50}}\right)\right)$. The scale parameter $\alpha_1 = \alpha_{50}$ remains the same. This results in state vector $(\mu_1, \alpha_1) = (0.014075, 0.001688)$ or equivalently $(m_1, v_1) = (0.01505, 0.14388)$. Here, consistency in transformation means that the fifty year maximum loading distribution is the resulting distribution of applying the one year loading fifty times independently, under assumption of constant resistance.

In the following, functions for representation of time effects are introduced. Although relying on realistic characteristics, the following time effect modeling is assumed for illustration purposes and does not claim to be appropriate in general. More versions of processes representing non-stationarity and degradation of structures were analyzed in Li et al. (2015b). Our model allows to use any of them. Time effects are introduced, in continuous time for $t > 0$ in years or in the period sense for the corresponding time interval $I_j \in \mathbb{I}$ of one year length, containing t and following the principle of safety, as:

$$\text{Non-stationary Loadings: } \theta_s(t) = 1.005^t \to \theta_s^I(I_j) = 1.005^j$$

$$\text{Post-Curing: } \theta_c(t) = 1.1 - \frac{0.1}{\sqrt{1+t^2}} \to \theta_c^I(I_j) = 1.1 - \frac{0.1}{\sqrt{1+(j-1)^2}}$$

Deterioration: $\theta_r(t) = 1 - 0.00012t^2 \rightarrow \theta_r^I(I_j) = 1 - 0.00012j^2$

As the initial distribution R_0 is assumed to be the same as that of R above, the heat map remains the same as in Figure 8.1a. It is shown in Figure 8.1c, where this time the states of the system $w(I_j)$ corresponding to a time period $I_j, j = 1, 10, 20, 30, 40, 50$ are marked and the initial state is given by

$$\begin{aligned} w(I_1) &= \left(\mu(1)\frac{\theta_s^I(I_1)}{\theta_c^I(I_1)\theta_r^I(I_1)}, \alpha(1)\frac{\theta_s^I(I_1)}{\theta_c^I(I_1)\theta_r^I(I_1)}\right) \\ &= \left(\mu(1)\frac{\theta_s(1)}{\theta_c(0)\theta_r(1)}, \alpha(1)\frac{\theta_s(1)}{\theta_c(0)\theta_r(1)}\right) \\ &= \left(0.01408 \cdot \frac{1.005}{1.0 \cdot 0.9999}, 0.00169 \cdot \frac{1.005}{1.0 \cdot 0.9999}\right) \\ &= (0.01415, 0.00170) \end{aligned}$$

represented as $(m(I_1), v(I_1)) = (0.015127, 0.14388)$. To judge the development of the failure probability, also the 50-year benchmark was scaled down to one year. The new benchmark is 1.3×10^{-6}, where good reason for this choice can be verified by $1-(1-1.3\times10^{-6})^{50} = 6.5\times10^{-5}$. Important to know is that although there might be instabilities in the supporting lines by interpolation, the real values and also the colors of the heat map are not derived by interpolation, but by direct computation of the integral in (7.1) (or (7.6)) instead. Due to the factor model, all time effects, except for discretization, result in a constant coefficient of variation and therefore only in horizontal shifts with respect to time, on this map. The heat map allows to directly see which parameters lead to violation of benchmarks such as 1.3×10^{-6}. At the same time, robustness of the result can be assessed by distance of the state to the benchmark. To also derive the cause and controlling opportunities of time effects, the bar chart in Figure 8.1d decomposes the theoretical mean value of the loading variable into its single components.

Fortunately, this is straightforward in the factor model. From a practical point of view, this representation yields good support for dominant effect derivation. Combining heat map and bar chart results in a powerful toolbox for extensive time-dependent reliability analysis. For instance, considering the state $(0.01787, 0.14388)$ corresponding to I_{30} and $p_f = 7.85 \times 10^{-7}$, one sees that small increases in m will result in a benchmark violation and the effect of changing the coefficient of variation. Together with the bar chart,

the reasons for benchmark violation at I_{40} with 1.34×10^{-5} are found quickly. It is exactly the time where the deterioration effect becomes dominant. This suggests inspections or necessary actions to reduce the expected deterioration. To show how such analysis could be extended, also the state which suits reliability after a column failure, thus in a 2-span slab (compare Figure 6.14b) with a doubled length at one span, was added.

Next, the survival effect is considered (using (7.6)), which changed the failure probabilities as follows:

$$I_{10}: \; p_f = 1.06 \cdot 10^{-8} \text{ becomes } p_f = 1.059 \cdot 10^{-8}$$

$$I_{30}: \; p_f = 7.90 \cdot 10^{-7} \text{ becomes } p_f = 7.89 \cdot 10^{-7}$$

$$I_{50}: \; p_f = 3.29 \cdot 10^{-4} \text{ becomes } p_f = 3.27 \cdot 10^{-4}$$

In all cases, as expected, one has a reduction in failure probability where the effect was found to be negligible because of the very small failure probabilities. Nevertheless, it was thereby verified that the survival effect does not need to be considered. More information and thus a stronger survival effect follows when the dominant uncertainty comes from the loading variable or when the failure probability is higher in general. For demonstration, the loading distribution was fixed in the last regarded year and the resistance was varied over time. It was chosen as decreasing, constant or increasing over time, always ending up in the same parameter state $(m_S, v_S) = (0.035, 0.1)$ at I_{50} for best comparison. The rate of relative change was chosen as 0.267% per year. The results are shown in Table 8.1 where it remains to perform an extensive case study which is beyond the scope of this work. In Table 8.1, it is shown how survival effects can be highly relevant. For example a decrease in failure probability of about 60% percent in period I_{50} in case of constant resistance was achieved. Furthermore, the high relevance of time effect history was discovered, especially when the loading effect decreases, or respectively the resistance effect increases, over time.

8.1.2 Analysis of Complex Networks[3]

This example is dedicated to the analysis of complex networks. It is shown that an adaption to analysis of highly complex networks is possible, profiting from the new representation and the computational efficiency of the

[3]Based on Blandfort et al. (2021).

Table 8.1: Blandfort et al. (2021). Comparison of the survival effect under different time effects on loading variable $(S_t)_{t\geq 0}$ in the 3-slab example. For demonstration of the survival effect, the failure probability over time with respect to three different scenarios is shown; increasing, constant and decreasing demands $(S_t)_{t\geq 0}$. The relative rate of change per period was assumed as $\theta_s(I_j) = (1.0 + 0.267\%)^j$ where initial distributions where set to fulfill $(m_S, v_S) = (0.035, 0.1)$ at I_{50} in every case. This way, period I_{50} allows to compare the different historical progress in the three cases, where constant R_t demonstrates the plain survival effect. Note that instead of decreasing (increasing) demands $(S_t)_{t\geq 0}$ one can equivalently consider the effect of increasing (decreasing) capacities $(R_t)_{t\geq 0}$.

Case	Probability of failure p_f^* $[\times 10^{-3}]$					
Time period	I_1	I_{10}	I_{20}	I_{30}	I_{40}	I_{50}
Increasing S_t	1.12	1.53(1.70)	2.15(2.67)	2.97(4.13)	4.04(6.33)	5.40(9.56)
Constant S_t	9.56	6.81	5.51	4.74	4.21	3.81
Decreasing S_t	59.31	20.97(43.56)	9.93(30.44)	5.52(20.95)	3.22(14.21)	1.91(9.51)
$*$ p_f given with and without survival effect. The unconditional p_f was added in braces.						

approach, also for a completely different type of structure. In Zuev et al. (2015) it was shown how to use Subset Simulation for analyzing the probability of satisfying maximum flow demands μ^* in two network topology types with stochastic link capacities. This approach was followed here, showing how network reliability can be considered in the model too, identifying maximum flow demands as S and network maximum flow capacity as R. For details, the reader is referred to Zuev et al. (2015). A small-network ring model $\otimes(n, k)$ with $n = 25$ and $k = 2$ is considered for demonstration. However, note that the procedure remains the same for arbitrary numbers of knots and edges so that this is an example for analysis of a structure that can take any dimension and become extremely computational demanding with respect to the evaluation of its capacity. A realization of a similar network is shown in Figure 8.2, where for illustration purposes, $n = 10$ and $k = 3$ were chosen. Link capacities were assumed independent and standard uniformly distributed. Dependencies could also be modeled by transformation of the variable space (cf. Rosenblatt Hohenbichler and Rackwitz (1981) or Nataf transformation (Der Kiureghian and Liu (1986))).

The corresponding limit state function requires to choose a source-sink pair of nodes (a,b), given by any two distinguished nodes of the network. The

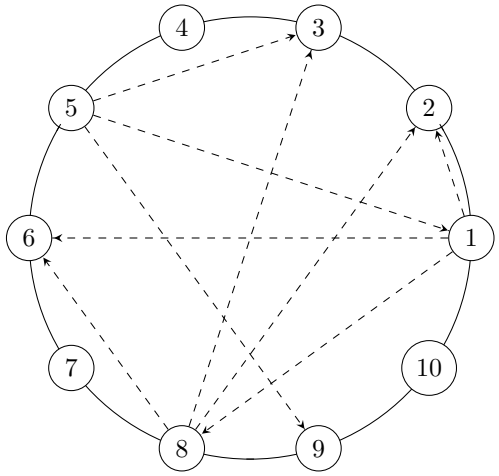


Figure 8.2: Blandfort et al. (2021). An exemplary small world ring network with $n = 10$ nodes and $k = 3$ additional edges per node. Only additional edges starting at $1, 5$ and 8 were drawn for illustration. Note that in the regular ring model edges are undirected links in contrast to the directed additional ones. It is assumed that additional edges are outgoing ones always.

object of interest is, whether the maximum flow demands from the source a to the sink b can be satisfied by the capacity of the network, represented by the limit state function

$$g(\mathbf{X}, (a, b), \mu^*) = \text{maxflow}(\mathbf{X},(\text{a,b})) < \mu^*, \tag{8.1}$$

where in the capacity-demand model $R \triangleq \text{maxflow}(\mathbf{X},(\text{a,b}))$ and $S \triangleq \mu^*$. The maximum flow capacity was computed by the Ford-Fulkerson algorithm Ford and Fulkerson (2009). All methods utilized in the 3-span slab example can be applied as well. Figure 8.3 shows the heat map for one randomly produced network. Averaging over several network realizations and source-sink pairs can in general allow for more profound claims. Again after one run of Subset Simulation (Algorithm 3), the failure probability with respect to any deterministic maximum flow demand or a distribution thereof can be evaluated by a one-dimensional integration. For illustration, time effects were added to the network model, where the progress over time for $\mathbb{I} = \{I_1, I_{10}, I_{20}..., I_{100}\}$ is also shown in Figure 8.3. This only serves for illustration of the parameter state model and is not based on any real world

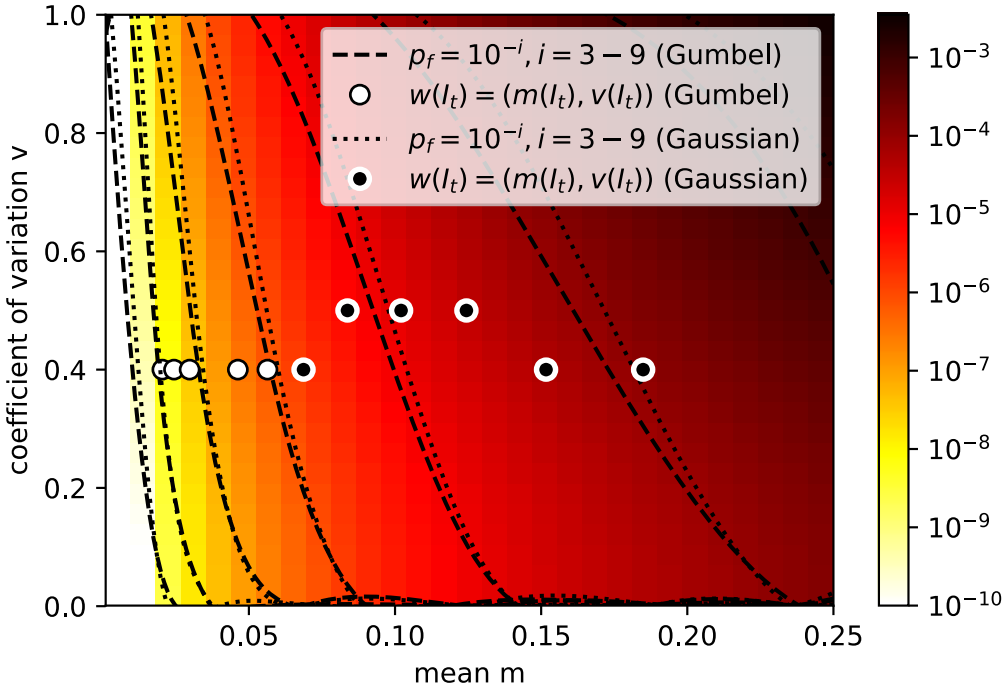


Figure 8.3: Blandfort et al. (2021). Visualization of time-dependent reliability of a small world ring network realization (cf. Figure 8.2) in the parameter state model. The loading variable is represented by a Gaussian distribution after the 40-th period I_{40}. For better comparison, no new heat map was drawn. Instead only the supporting lines as well as the corresponding parameter states belonging to a Gaussian distributed loading variable are shown.

evidence. Starting with $(m, v) = (0.02, 0.4)$, the average maximum flow demand m is then increased by 2% per period. Additionally, an absolute value of 0.01 was added to m after period I_{30} while the coefficient of variation v was kept constant. At the end of period I_{40}, the maximum flow demand distribution was changed to a Gaussian one, resulting in a new heat map afterwards. This new heat map was not drawn but instead indicated by additional supportive lines in Figure 8.3 so that it displays the robustness of reliability with regard to the selection of a distribution type of the maximum flow demand particularly well. Furthermore, the coefficient of variation was increased to 0.5 at period I_{60} and decreased back to its initial value at I_{90}. The corresponding failure probabilities can also be calculated without significant efforts in all cases. Note that the Gaussian distribution is also in the location-scale family, fulfilling (G2). As discretization is not relevant here, this suffices the demands, not needing property (G1). There are many opportunities to manipulate the model without significant additional computational demands, possibly allowing to answer several interesting questions in various fields such as transportation networks, web infrastructure or any other field where analysis of complex networks with high computational demand is necessary.

8.1.3 Lossy Transmission Line

As an example where FORM and SORM have been shown to fail, we consider a reliability problem given in Bourinet (2018) for application of the parameter state model. A lossy transmission line is analyzed, evaluating the probability of exceeding a given current magnitude level. This example also demonstrates the opportunity to switch the roles of R and S in the parameter state model (compare Section 7.5) Here, the limit state function is given by

$$g(\mathbf{X}) = I_{cr} - I(\mathbf{X})$$

where $I_{cr} = 1.5 \times 10^{-4} A$ is set fixed deterministic and $I(\mathbf{X})$ is explicitly given but includes $d = 11$ stochastic variables and implies a highly non-linear geometry of the limit state function (compare Bourinet (2018)). Note that the capacity is given by a deterministic value and complexity of the model comes from the structure of S. To make this example suitable to our model, we derive the cdf of $-I(\mathbf{X})$ by Subset Simulation. We proceeded by Algorithm 3 with $p_0 = 0.1$, $N = 2500$ and evaluation up to exceeding probabilities as low

as $p_{\min} = 1 \times 10^{-7}$, resulting in a total of 17500 limit state evaluations under neglecting re-use of seeds in Subset Simulation. The coefficient of variation was found as low as 0.184 (Bourinet (2018)). A single reliability evaluation of FORM and SORM requires around $10,000$ limit state evaluation and both approaches were even shown to be highly biased (Bourinet (2018)), hence stochastic methods are required. Repetitive calculations for different constellations by FORM and SORM would thus already result in more limit state evaluations than evaluation by Subset Simulation with $N = 2500$ and up to probabilities as low as 1×10^{-7} for consideration of $n = 2$ constellations. Subset Simulation allows to evaluate reliability for many different distributions of I_{cr}. However, for a deterministic I_{cr} as given in the model, we do not need an integration and only need to consider the conditional failure probability at the given value, that is directly available by the cdf. To get some information on uncertainty additionally, thereby receiving a heat map, we artificially assumed a Gaussian distribution of I_{cr}. The heat map is shown in Figure 8.4. Considering a deterministic I_{cr} then just refers to the abscissa in the illustration, where the coefficient of variation approaches zero. Also note that deterministic values suffice our demands for linear transformation so that we can also analyze effects of a linear transformations on $I(\mathbf{X})$ directly. If we now assume demands $I(\mathbf{X})$ increase by 2% per time unit, setting up a time-dependent model for illustration of the parameter state model in this case, results in estimated failure probabilities

$$I_1 : \; p_f = 2.12 \cdot 10^{-4}$$

$$I_5 : \; p_f = 6.07 \cdot 10^{-4}$$

$$I_{10} : \; p_f = 1.47 \cdot 10^{-3}$$

$$I_{20} : \; p_f = 6.58 \cdot 10^{-3}$$

over time. It is important to also directly evaluate some failure probabilities to validate the accuracy of the supporting lines for better readability of the heat map which have been constructed by smoothing splines, since too smooth results might be look appealing but disrespect accurate reliability evaluation at some points. Particularly note the heat map representation for a switched role for R and S .

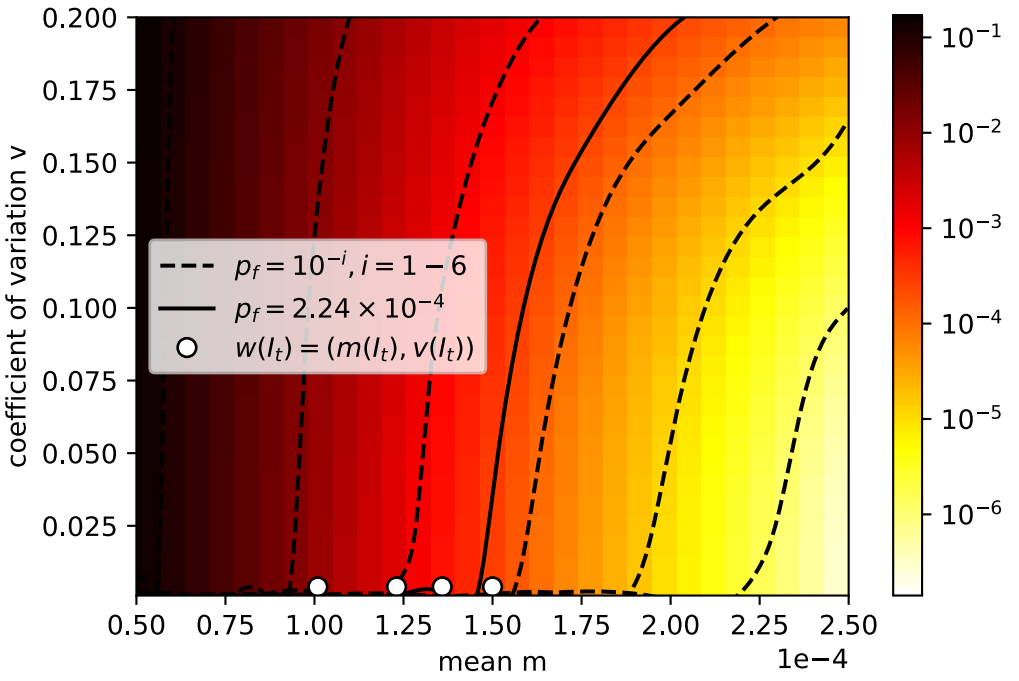


Figure 8.4: Analysis of a lossy transmission line. The role of capacity and demand is switched in this example. We thereby condition on the capacity variable $S = I_{cr}$. To extend the information beyond deterministic values and to consider also stochastic capacities, we artificially assumed a Gaussian distribution for the capacity I_{cr}. Having a deterministic value $I_{cr} = 1.5 \times 10^{-4}$ for the current magnitude level not to be exceeded, which represents the capacity in our model, the corresponding state is on the abscissa. Additional, we added a hypothetic time-dependent scenario, assuming $I(\mathbf{X})$ increases by 2% per time unit.

8.2 Subset Simulation Interpolation

For analysis of SuSI, we can use the exact same visualization technique as for the parameter state model, as both methods are based on computation of failure probabilities with respect to the distribution of a single stochastic variable, available by interpolation of SuS and SuSI results, respectively. We consider three examples, a continuation of the extensive analysis of our illustrational slab example (Section 6.2.2 and Section 8.1.1), the network example as studied in Section 8.1.2 and a two degree of freedom damped oscillator as given in Section 6.1.1.

8.2.1 Time-Dependent Model

In this section, we continue with the analysis of a 3-span slab as given in Section 6.2.2 and Section 8.1.1. The example well shows how, in contrast to the parameter state model, one can change the distribution of any variable of the stochastic reliability model, which fulfills the milder assumptions of monotonicity and independence, in an arbitrary way instead of being restricted to capacity or demand variable. This allows to consider much more sophisticated time-dependent reliability models. Note that another example on application of SuSI for time-dependent models was already given in Section 6.2, utilizing the opportunity to consider artificial variables for modeling the model dynamics and thereby allowing to analyze time-dependent effects on several model variables.

Here, we pick the cross sectional area of the steel (A_s) as dynamic variable and investigate the effect of changes in its distribution on reliability. A practical example, where such considerations might be particularly useful, are structures which are exposed to marine environments and suffer from corrosion. Then $A_s(t)$ is a stochastic variable that evolves with respect to time, results in a different probability distribution for differing time instants, is exposed to high uncertainties, and most likely yields the dominant impact on time-dependent reliability of the structure. To study the effects on reliability by corrosion, we assume a scenario where $A_s(t)$ is lognormal distributed with mean value $m(I_t) = 5.24 \cdot 10^{-4} - t \cdot 5 \cdot 10^{-6}$ and coefficient of variation $v(I_t) = 0.01 \cdot (1.075)^t$ in year I_t, $t \in \mathbb{N}$. As in the parameter state model, we also refer to the corresponding state at time I_t by $w(I_t)$. The resulting visualization is given in Figure 8.5. If we use the failure probability benchmark $1.3 \cdot 10^{-6}$, it is clearly visible which distributions become critical

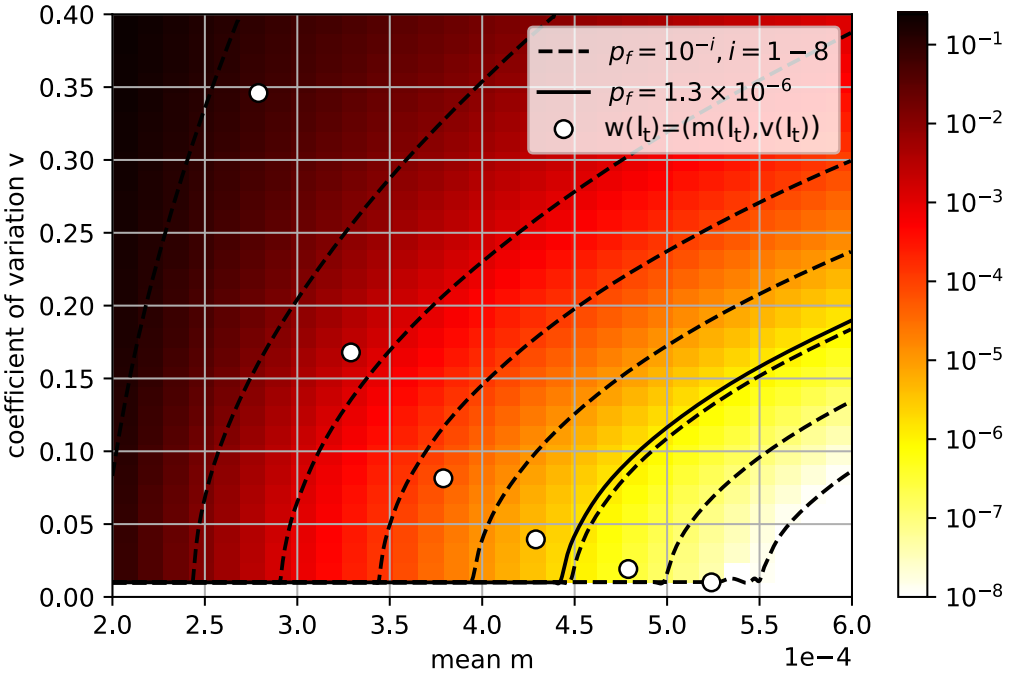


Figure 8.5: Reliability with respect to variation of the cross sectional area of the steel A_s. Assuming A_s as the dominant time effect, reliability over time is shown for years $I_1, I_{10}, I_{20}, I_{30}, I_{40}$ and I_{50}. $A_s(I_t)$ was assumed lognormal distributed at all times.

ones and also that such a distribution is assumed for year I_{20} in our scenario. Because corrosion depends on many factors, uncertainty takes an important role in such an analysis. Beyond the problematic of predicting its distribution, one is also often interested in creating good plans for inspection of such structures. Consequently, such analysis might help to trace when uncertainties become unbearable, supporting the engineer in making appropriate and cost-optimized inspection plans.

8.2.2 Network Example

Also analyzing the network as given in Section 8.1.2, we now show how SuSI can consider alternative model dynamics for complex networks. As the studied network topology can take an arbitrary size and large networks are an important field of application for reliability analysis, this example gives a particularly good inspiration for extraordinary opportunities. Here, we again consider a small-network ring model $\otimes(n, k)$ with number of knots n and number of additional edges per node k, resulting in a total of $k^* = n \cdot (k + 1)$ edges. Again a source-sink pair is considered and the maximum flow is examined, resulting in limit state function (8.1). The idea is to let the dynamic variable $\mathbf{X}_k$ relate to specific network properties, modifying the topology or distribution of capacities of the edges. Of course, we could set the distribution of a single edge as the dynamic variable. However, this would give minor information only in such a network topology with many nodes and edges. Nevertheless, if considering more sophisticated versions of networks where distributions of the edges vary and outstandingly important edges exist, e.g. a highway representing an edge in a traffic network, this could be of high interest. Also, we can multiply an artificial variable to the capacities of the single realizations of an arbitrary set of the edges of the network, decreasing or increasing the maximum flow. Anyhow, we consider a very special case that can be considered in such a network topology and is not as obvious as the discussed opportunities. Here, the dynamic variable $\mathbf{X}_k$ is considered as the number of edges that is removed from the network, or equivalently the number of edges which take zero capacity. In reality, this would typically refer to failure of these single node connections.

The procedure is the following. First, we order all k^* edges of the network at random. According to this ordering, we then remove all edges e_1, e_2,...e_{y_1} for some $y_1 \in \{1, ..., k^* - 1\}$, beginning with a high number of edges removed, e.g. $y_1 \approx 0.9k^*$, for estimation of a failure probability. Having only a few

edges left, the capacity of the network maxflow($\mathbf{X}$,(a,b)) becomes low. This defines the first limit state function, considering the maximum flow from source to sink in the corresponding network with $k^* - \mathbf{x}_k$ edges. Now SuSI is applied, either starting with one crude Monte Carlo simulation or using Subset Simulation to achieve a high enough failure probability.

Then, the number of edges is successively increased, only removing edges e_1, e_2,...e_{y_i}, $i = 1, 2, ...$ up to edges $y_1 > y_2 > ...$ until SuSI terminates, reaching a failure probability below the maximum error or $\mathbf{x}_k = 0$ which defines the original network topology without removing any edges. By interpolation according to SuSI, we then receive the conditional failure probability of the network with respect to the number of removed edges. Keep in mind that this result is based on a fixed source-sink pair as well as an ordering of the edges to keep monotonicity with respect to the limit state. If however, we want to consider more scenarios or average over a specific distribution of edge orderings (e.g. uniform), then we can just repetitively apply the procedure and average over all results.

We considered two cases for demonstration, $\otimes(20, 15)$ with $\mu^* = 1$ as well as $\otimes(25, 20)$ with $\mu^* = 5$. The resulting visualizations are shown in Figure 8.6 and Figure 8.7, respectively. In this example, we have shown how networks can be analyzed in detail and at a low cost and how to present the result by our visualization technique.

8.2.3 Two Degree of Freedom Damped Oscillator

In this section, we want to continue with the example studied in Section 6.1, considering a two degree of freedom primary/secondary damped oscillator (see Der Kiureghian and De Stefano (1991); Bourinet et al. (2011)) with limit state function (6.1) and stochastic properties as in Table 6.4. Here, we want to point out how the results given in Section 6.1 enable us to create extensive reliability analysis at a low cost. The resulting visualizations are given in Figure 8.8. All cases ($\mathbf{X}_k = \zeta_p, \zeta_s, F_s, S_0$) provide an approximately correct estimation of the failure probability in the static setting, being close to the solid curve, which marks the states corresponding to estimated failure probabilities equal to the true failure probability. Having a clear picture on robustness of the failure probability with respect to the distributions of many random variables allows to judge importance of individual stochastic variables in the model and possibly to interpret and understand reliability of the problem more globally. Note, it is important to consider only dy-

Figure 8.6: The impact of failed edges on the network reliability with respect to network topology $\otimes(20, 15)$ and maximum flow demand $\mu^* = 1$. The corresponding heat map with respect to lognormally distributed $\mathbf{X}_k$ = 'Number of edges removed' was drawn.

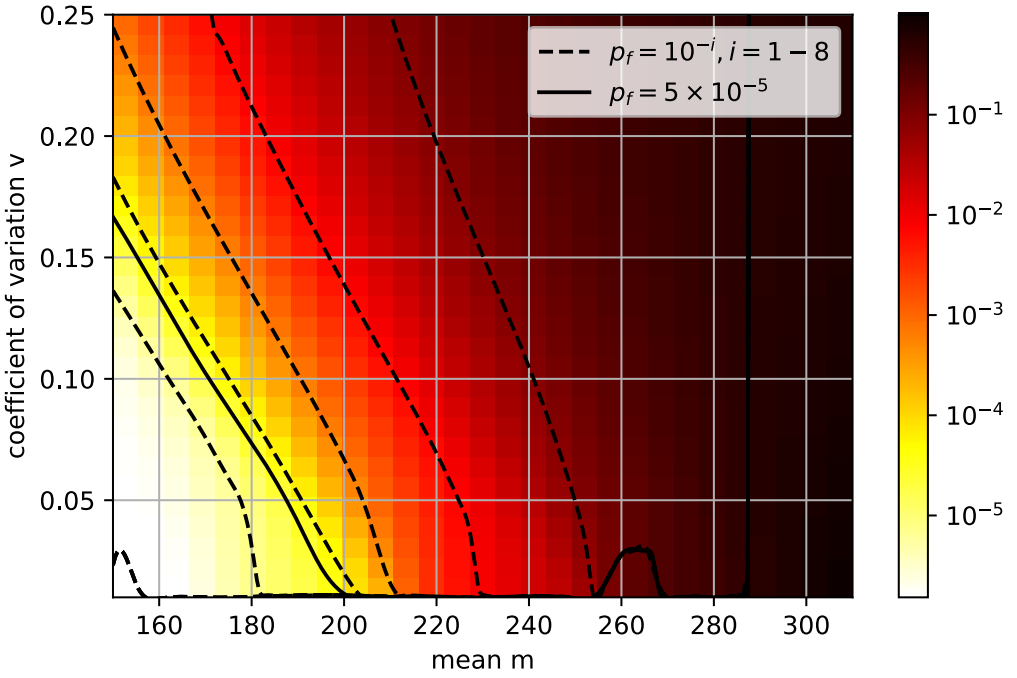
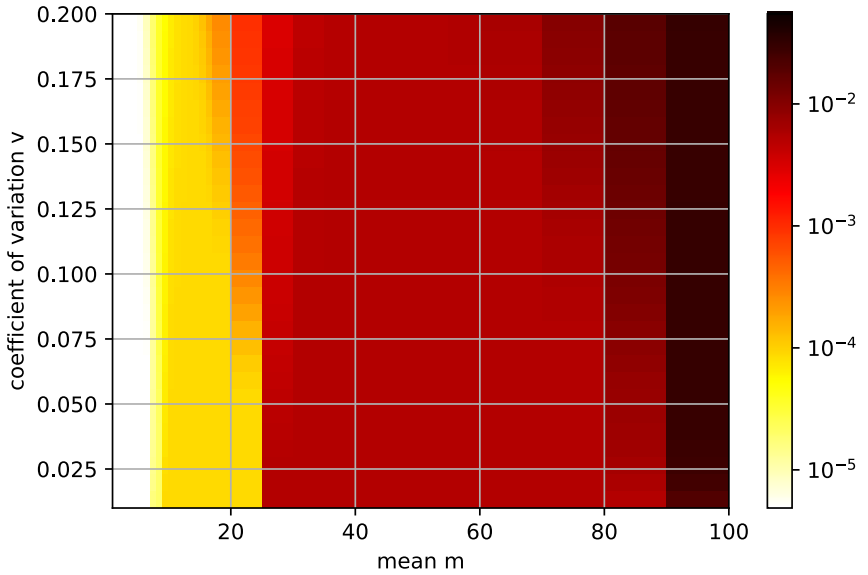
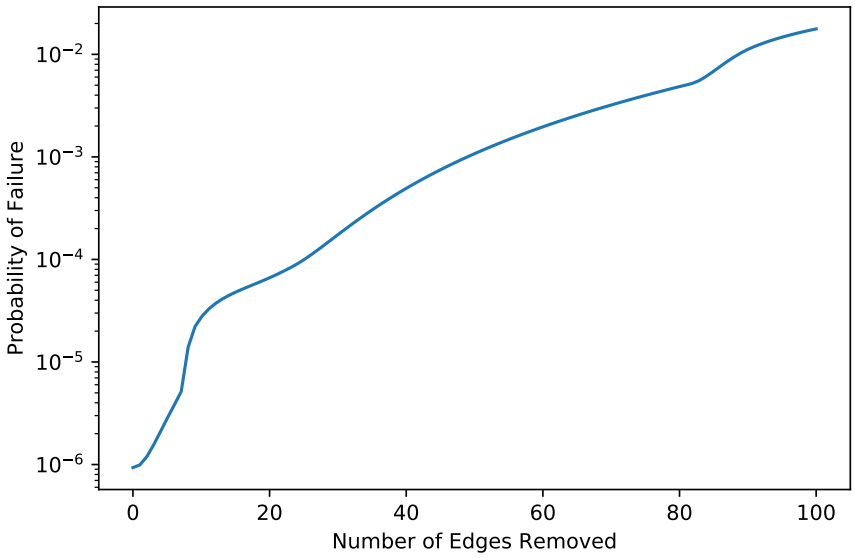


Figure 8.7: The impact of failed edges on the network reliability with respect to network topology $\otimes(25, 20)$ and maximum flow demand $\mu^* = 5$.

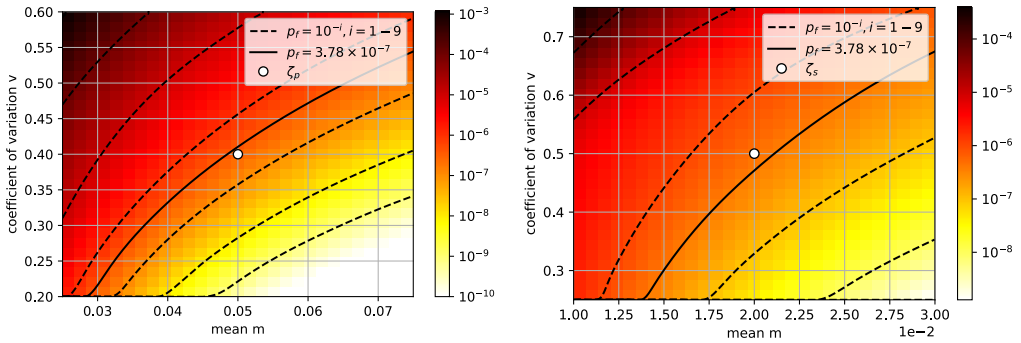


(a) Heat map with respect to lognormally distributed $\mathbf{X}_k$ = Number of edges removed.



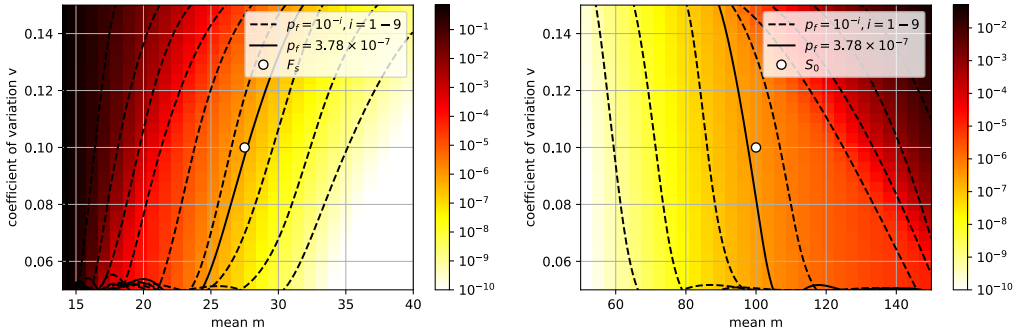
(b) Conditional failure probability with respect to the numbers of non-available edges.

Figure 8.8: Extensive Study: Considering sensitivity of the failure probability with respect to several random variables of the stochastic model of a two degree of freedom damped oscillator. The solid curve marks the states with estimated failure probabilities equal to the true failure probability. Optimally, all considered dynamic variables should lead to a state that hits this curve perfectly.



(a) $\mathbf{X}_k = \zeta_p$

(b) $\mathbf{X}_k = \zeta_s$



(c) $\mathbf{X}_k = F_s$

(d) $\mathbf{X}_k = S_0$

namic variables which fulfill the necessary assumptions (monotonicity and independence) for application of SuSI. Otherwise, results might be severely biased, in particular if monotonicity is not given. On the other hand, we might also get such extremely biased results by calculation with SuS, where using SuSI on several variables allows to possibly identify such shortcomings by discrepancies in results of individual calculations.

8.3 Conclusion

In this thesis, computational efforts were reduced and comprehension was increased, by mapping all dynamics of the stochastic model to a single variable, in both approaches, parameter state model and SuSI. In this chapter, we showed how this allows for a dynamic reliability analysis by analyzing changes in the distribution of this single variable, creating the opportunity for the presented visualization technique. The parameter state model and SuSI, in combination with the new visualization techniques offer many opportunities to extensively analyze several complex civil engineering structures in an efficient way. For demonstration, we analyzed some civil engineering structures and presented their corresponding heat maps, in this chapter. The heat maps allowed for a novel and holistic perspective on model uncertainties and reliability at the same time.

Chapter 9

Conclusions and Outlook for Future Research

In this thesis, we have presented a novel algorithm 'Subset Simulation Interpolation' (SuSI) for efficient reliability estimation under model-dynamics, extensively studied algorithmic properties of ordinary Subset Simulation, demonstrated how to greatly benefit from utilization of mathematical properties in specific time-dependent models and applied new visualization techniques that perfectly match the presented approaches.

A main idea throughout this thesis was the linking of modeling and efficient reliability evaluation, resulting in SuSI and the parameter state model. As stated in the introduction, we were required to demand restrictions on the considered dynamic model, but thereby gained the ability to introduce algorithms that solve the defined model with high efficiency. Moreover, many models which seemed to be of a different kind, could be converted to such dynamic models without losing much accuracy. If the computational demand of the original model is infeasible, this even is an inevitable approach. In this regard, we directly faced the challenge of finding a good balance between computational effort and accuracy. Although a loss in accuracy can often be avoided by sophisticated modeling, we generally trade accuracy for efficiency.

Furthermore and particularly important in analysis of complex structures, we dedicated ourselves to the understanding of results provided by reliability analysis with Subset Simulation. For this purpose, we extended the statistical analysis of ordinary Subset Simulation, leading to new findings and an improvement in understanding of the algorithm, and demonstrated how results by SuSI and the parameter state model can be visualized in an informative

way.

As a detailed list of all contributions was given in Section 2.3, we particularly highlight the (practical) consequences of this thesis for reliability analysis and give an outlook for future research in the following.

Chapter 4: Ordinary Subset Simulation. First, there are consequences for application of ordinary Subset Simulation. It was shown that the efficiency of Subset Simulation depends mainly on the Markov chain length instead of p_0. Choosing chain lengths such as N_l between 10 and 20 yields a similar equivalent efficiency, independent of p_0 for $p_0 \geq 0.1$ and is thus recommended to set as a fixed value instead of an adaptive one. Furthermore, implementation of the algorithm has been modified to suit theory better. Seeds of the subsets that hit the threshold, and those correlated strongly to these, have not been considered for the next subset. As a consequence, the negative bias in current implementations of Subset Simulation was removed. At first, this resulted in a higher positive Subset Simulation bias, but allowed for a bias correction by the explicitly derived result in statistical analysis, as simulation results were now more in line with theoretical results. In total, the positive bias in Subset Simulation was thereby reduced, also resulting in a slightly lower variance under constant coefficient of variation. A stochastic interpretation for extended understanding of Subset Simulation was given, allowing for an explicit bias correction formula as well as for reconsidering distributions of intermediate probabilities at the subset level and proving independence of these (for successive subsets) under weak assumptions. Evidence by simulations was found under appropriate implementations of Subset Simulation.

For future study, it would be interesting to search for examples that lead to a negative bias of Subset Simulation when following the threshold selection present in current implementations. Based on the given stochastic interpretation and the demonstrated sample dependencies with respect to p_0, another formula for the coefficient of variation could be derived, where a first step towards estimating within subset sample dependencies, given p_0, was made. The opportunity to also use higher p_0 values for simulation at the same efficiency could allow for novel modifications of the algorithm that have not yet been explored due to the restriction to small p_0. For example, SuSI benefits a lot from this finding.

Chapter 5 and Chapter 6: Subset Simulation Interpolation. Second, the novel algorithm Subset Simulation Interpolation (SuSI) allows for many new opportunities for efficient reliability evaluation and also for validation of results by ordinary Subset Simulation. Consequently, new extensive case studies with respect to dynamic models can be carried out in cases that might have come at a too high computational demand formerly. In particular, we achieve better tractability of accuracy for considered model constellations, because the approximation error by interpolation can be bounded and does, in contrast to e.g. Taylor expansion for unknown functions, only include systematical errors that allow for secure boundaries on the reliability estimation. This becomes even more relevant for substantially differing model constellations with respect to the dynamic variable. Also, instead of considering fixed constellations, information according to the result by SuSI allows for analysis of any distribution of the dynamic variable. We do not need to choose settings in advance, since results for different distributions can build up on the same SuSI simulation.

Beyond the application on dynamic models, such as time-dependent reliability analysis, that can naturally be set up with SuSI, we also explored the possibility for validation of results by ordinary SuS. To our best believe, it does not require any relevant model assumptions, as an artificial variable is introduced that should generally fulfill the demands on the given dynamic model. This is an observation with extremely high potential, since both algorithms perform best at the same reliability problems, which in turn allows cheap backup tests or a more robust algorithm on its own, alternating to SuS. The interesting point is the source of systematically wrong results in SuS, which appear for some geometries of the limit state functions. This makes SuSI a good candidate for solving such issues, because it can alter the geometry of the limit state, as desired, and allows to set up many ways for evaluation of the failure probability. A promising way to alter limit state functions for robust reliability analysis was presented.

The novel algorithm SuSI offers a new approach on reliability analysis. In this thesis, we have already been able to elaborate a good theoretical basis and many possible applications. However, there are so many possibilities for further development that the future study and extension of the algorithm could become an important research topic. For this purpose, an implemen-

tation of the algorithm is also publicly available at GitHub[1]. We are looking forward for collaborations and future extensions.

In particular, further development of methods such as a robust reliability analysis, where the limit state function is artificially altered so that the algorithm provides correct results, could be interesting. Having the opportunity to control the geometry of the limit state function while keeping the same failure probability, allows to create many backup tests for SuS results as well as stand-alone approaches which can possibly be backed up by a theoretical study better. We could, for example, examine if we can define the successive limit states in SuSI such that the structure of these guarantees to find the most important regions of the sample space with high probability, independent of the original limit state function. The robustness of such a target-oriented modification of the geometry of the limit state function by SuSI should then be validated mathematically. Also applications in network reliability and resilience are a promising field of study. Additionally, utilizing other post-processing approaches for ordinary Subset Simulation for SuSI could provide even more information on model-dynamics. Furthermore, examining the usage of regression instead of interpolation to get a continuous failure probability function with respect to the dynamic variable would yield a practically unbiased result, avoiding a potential drawback of the algorithm and removing the last bit of uncertainty by numerical methods, beyond the one on top of stochastic variation in ordinary SuS.

According to the manifold fields of application of ordinary SuS, many opportunities for dynamic model representations as the ones presented, where SuSI yields an efficient evaluation, can most likely be explored. The decision about selecting the dynamic variable either as an artificially introduced one, or dependent on relevance of variables of the given model, is of course up to the specific experts with experience in the corresponding field of application.

Chapter 7: Efficient Time-Dependent Reliability Evaluation by a Parameter State Model. Third, we considered a specific case of a time-dependent model. The capacity-demand setting allows to efficiently compute reliability with respect to several different distributions of the loading variable. If this variable belongs to the location-scale family, it is even possible to apply such transformations to evaluate time-dependent effects on the capac-

[1] https://github.com/FBlandfort/Subset-Simulation-Interpolation.

ity, without needing additional reliability evaluations, directly receiving the failure probability by a computational negligible one-dimensional integration. On top of this finding, a formula for reliability evaluation under consideration of past survival was derived. In general, this efficient framework allows to conduct extensive case studies at a feasible cost and shows how it can be beneficial to favor specific types of models over others, if accuracy is kept high enough under an appropriate model conversion.

In future work, an interesting extension of the model would be the study of systems of such structures, where the single structures are interacting with each other. For example considering several networks, each corresponding to a capacity-demand model, setting up mechanisms which are triggered when specific thresholds are exceeded in individual networks. The mechanism could then produce an increased demand on other networks, such as a breakdown of traffic could produce an increased demand on web services, or just on other traffic networks that managed to keep their flow capacity.

Chapter 8: Informative Visualization. Fourth, we showed how to apply the introduced approaches for efficient reliability evaluation and how to present results in a valuable way. The illustration by a heat map provides an extensively informative view on robustness and uncertainty or also on the dynamics with respect to one variable of the model and is therefore perfectly suitable for the developed approaches. Additionally, if the clear cut structure of the presented parameter state model can be utilized, time-dependent effects can be visualized in detail by a bar chart representation. Both, heat maps and bar charts aim at and achieving a holistic perspective on reliability analysis, making the effect of uncertainties or model dynamics with respect to a stochastic variable of the model directly visible.

Visualization techniques and new perspectives on reliability issues are important in analysis of complex structures. In perception and presentation of results, complexity of any level has to be broken down to a small amount of information that can be well understood. Also uncertainties or unclear parts of the model should be directly visible in the presentation. An extension, penalizing small distances to a reliability benchmark instead of just considering being below or above a benchmark, could be an interesting topic for future considerations. Risk measures with meaningful utility functions, de-

pending on the specific application, could then allow for more sophisticated assessment of reliability analysis.

Bibliography

A. Abdollahi, M. A. Moghaddam, S. A. H. Monfared, M. Rashki, and Y. Li. Subset simulation method including fitness-based seed selection for reliability analysis. *Engineering with Computers*, pages 1–17, 2020.

D. A. Alvarez, F. Uribe, and J. E. Hurtado. Estimation of the lower and upper bounds on the probability of failure using subset simulation and random set theory. *Mechanical Systems and Signal Processing*, 100:782–801, 2018.

C. Andrieu-Renaud, B. Sudret, and M. Lemaire. The phi2 method: a way to compute time-variant reliability. *Reliability Engineering & System Safety*, 84(1):75–86, 2004.

B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*, volume 54. Siam, 1992.

S. Au. Reliability-based design sensitivity by efficient simulation. *Computers & Structures*, 83(14):1048–1061, 2005.

S. Au and J. L. Beck. First excursion probabilities for linear systems by very efficient importance sampling. *Probabilistic Engineering Mechanics*, 16(3): 193–207, 2001a.

S.-K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16 (4):263–277, 2001b.

S.-K. Au and E. Patelli. Rare event simulation in finite-infinite dimensional space. *Reliability Engineering & System Safety*, 148:67–77, 2016.

S.-K. Au and Y. Wang. *Engineering Risk Assessment with Subset Simulation*. John Wiley & Sons, 2014.

A. T. Beck and W. J. de Santana Gomes. A comparison of deterministic, reliability-based and risk-based structural optimization under uncertainty. *Probabilistic Engineering Mechanics*, 28:18–29, 2012.

F. Blandfort, C. Glock, J. Sass, S. Schwaar, and R. Sefrin. Subset simulation interpolation - a new approach to compute effects of model-dynamics in structural reliability. In M. Beer and E. Zio, editors, *Proceedings of the 29th European Safety and Reliability Conference (ESREL 2019)*, pages 1978–1986, Hannover, Germany, 2019a. ESRA.

F. Blandfort, C. Glock, J. Sass, S. Schwaar, and R. Sefrin. A parametric state space model for time-dependent reliability analysis. In D. Yurchenko and D. Proske, editors, *Proceedings of the 17th International Probabilistic Workshop (IPW 2019)*, pages 31–36, Edinburgh, UK, 2019b. Heriot Watt University.

F. Blandfort, C. Glock, J. Sass, S. Schwaar, and R. Sefrin. Efficient and comprehensive time-dependent reliability analysis of complex structures by a parameter state model. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 7(2):04021020, 2021.

J.-M. Bourinet. *Reliability Analysis and Optimal Design Under Uncertainty-Focus on Adaptive Surrogate-Based Approaches*. PhD thesis, 2018.

J.-M. Bourinet, F. Deheeger, and M. Lemaire. Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety*, 33(6):343–353, 2011.

K. Breitung. Asymptotic approximations for multinormal integrals. *Journal of Engineering Mechanics*, 110(3):357–366, 1984.

K. Breitung. On subsets and onions: Lost in outer space. In *Joint ICVRAM ISUMA Uncertainties Conference*, 2018.

K. Breitung. The geometry of limit state function graphs and subset simulation: Counterexamples. *Reliability Engineering & System Safety*, 182: 98–106, 2019.

C. Bucher. Asymptotic sampling for high-dimensional reliability analysis. *Probabilistic Engineering Mechanics*, 24(4):504–510, 2009.

J. Butland. A method of interpolating reasonable-shaped curves through any data. *Proc. Computer Graphics*, 80:409–422, 1980.

F. Cérou, P. Del Moral, F. Le Gland, and P. Lezaud. Genetic genealogical models in rare event analysis. *Latin American Journal of Probabability and Mathematical Statistics*, 1:181–203, 2006.

F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–808, 2012.

F. E. H. Chehade and R. Younes. Structural reliability software and calculation tools: a review. *Innovative Infrastructure Solutions*, 5(1):1–16, 2020.

J. Ching, S.-K. Au, and J. L. Beck. Reliability estimation for dynamical systems subject to stochastic excitation using subset simulation with splitting. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1557–1579, 2005.

F. Coggeshall. The arithmetic, geometric, and harmonic means. *The Quarterly Journal of Economics*, 1(1):83–86, 1886.

A. A. Czarnecki and A. S. Nowak. Time-variant reliability profiles for steel girder bridges. *Structural Safety*, 30(1):49–64, 2008.

M. De Angelis, E. Patelli, and M. Beer. Advanced line sampling for efficient robust reliability analysis. *Structural Safety*, 52:170–182, 2015.

L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Science & Business Media, 2007.

A. Der Kiureghian and T. Dakessian. Multiple design points in first and second-order reliability. *Structural Safety*, 20(1):37–49, 1998.

A. Der Kiureghian and M. De Stefano. Efficient algorithm for second-order reliability analysis. *Journal of Engineering Mechanics*, 117(12):2904–2923, 1991.

A. Der Kiureghian and P.-L. Liu. Structural reliability under incomplete probability information. *Journal of Engineering Mechanics*, 112(1):85–104, 1986.

DIN EN 1990: 2010-12. Eurocode: Basis of structural design. *Beuth Verlag GmbH Berlin*, German version EN 1990:2002 + A1:2005 + A1:2005/AC:2010, 2010.

O. Ditlevsen. Narrow reliability bounds for structural systems. *Journal of Structural Mechanics*, 7(4):453–472, 1979.

S. S. Dragomir. A survey on cauchy-bunyakovsky-schwarz type discrete inequalities. *Journal of Inequalities in Pure and Applied Mathematics*, 4(3): 1–142, 2003.

X. Du. Time-dependent mechanism reliability analysis with envelope functions and first-order approximation. *Journal of Mechanical Design*, 136 (8), 2014.

A. Elsheikh, S. Oladyshkin, W. Nowak, and M. Christie. Estimating the probability of co2 leakage using rare event simulation. In *ECMOR XIV-14th European Conference on the Mathematics of Oil Recovery*, volume 2014, pages 1–9. European Association of Geoscientists & Engineers, 2014.

S. Engelund and R. Rackwitz. A benchmark study on importance sampling techniques in structural reliability. *Structural Safety*, 12(4):255–276, 1993.

D.-Y. Fan. The distribution of the product of independent beta variables. *Communications in Statistics-Theory and Methods*, 20(12):4043–4052, 1991.

W. F. Ferger. The nature and use of the harmonic mean. *Journal of the American Statistical Association*, 26(173):36–40, 1931.

B. Fiessler, R. Rackwitz, and H.-J. Neumann. Quadratic limit states in structural reliability. *Journal of the Engineering Mechanics Division*, 105 (4):661–676, 1979.

L. R. Ford and D. R. Fulkerson. Maximal flow through a network. In *Classic Papers in Combinatorics*, pages 243–248. Springer, 2009.

F. N. Fritsch and J. Butland. A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific and Statistical Computing*, 5(2):300–304, 1984.

F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.

M. Fujita and R. Rackwitz. Updating first-and second-order reliability estimates by importance sampling. *Doboku Gakkai Ronbunshu*, 1988(392): 53–59, 1988.

B. Gaspar, A. Teixeira, and C. G. Soares. Assessment of the efficiency of kriging surrogate models for structural reliability analysis. *Probabilistic Engineering Mechanics*, 37:24–34, 2014.

S. Glowienka, A. Fischer, and M. Krau. Implementation of probabilistic methods in structural design. In H. Budelmann, A. Holst, and D. Proske, editors, *Proceedings of the 9th International Probabilistic Workshop*, pages 275–286. Technische Universität Braunschweig, 2011.

R. V. Grandhi and L. Wang. Higher-order failure probability calculation using nonlinear approximations. *Computer Methods in Applied Mechanics and Engineering*, 168(1-4):185–206, 1999.

J. Guo and X. Du. Reliability sensitivity analysis with random and interval variables. *International Journal for Numerical Methods in Engineering*, 78 (13):1585–1617, 2009.

C. A. Hall and W. W. Meyer. Optimal error bounds for cubic spline interpolation. *Journal of Approximation Theory*, 16(2):105–122, 1976.

A. M. Hasofer and N. C. Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics division*, 100(1):111–121, 1974.

M. Hohenbichler and R. Rackwitz. Non-normal dependent vectors in structural safety. *Journal of the Engineering Mechanics Division*, 107(6):1227–1238, 1981.

W.-C. Hsu and J. Ching. Evaluating small failure probabilities of multiple limit states by parallel subset simulation. *Probabilistic Engineering Mechanics*, 25(3):291–304, 2010.

H. T. Huynh. Accurate monotone cubic interpolation. *SIAM Journal on Numerical Analysis*, 30(1):57–100, 1993.

A. M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *University of Cambridge, Department of Engineering, Cambridge University Engineering Department, Trumpington*. Citeseer, 2005.

L. S. Katafygiotis and S. Cheung. Application of spherical subset simulation method and auxiliary domain method on a benchmark reliability study. *Structural Safety*, 29(3):194–207, 2007.

H.-J. Kim and D. Straub. Efficient computation of the lifetime reliability of deteriorating structures. In *Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13)*, 2019.

A. Kouassi, J.-M. Bourinet, S. Lalléchère, P. Bonnet, and M. Fogli. Reliability and sensitivity analysis of transmission lines in a probabilistic EMC context. *IEEE Transactions on Electromagnetic Compatibility*, 58(2):561–572, 2016.

P.-S. Koutsourelakis, H. J. Pradlwarter, and G. I. Schuëller. Reliability of structures in high dimensions, part i: algorithms and applications. *Probabilistic Engineering Mechanics*, 19(4):409–417, 2004.

H. Li, Z. Lü, and X. Yuan. Nataf transformation based point estimate method. *Chinese Science Bulletin*, 53(17):2586, 2008.

H.-S. Li and Z.-J. Cao. Matlab codes of subset simulation for reliability analysis and structural optimization. *Structural and Multidisciplinary Optimization*, 54(2):391–410, 2016.

H.-S. Li, Y.-Z. Ma, and Z. Cao. A generalized subset simulation approach for estimating small failure probabilities of multiple stochastic responses. *Computers & Structures*, 153:239–251, 2015a.

J. Li, J.-b. Chen, and W.-l. Fan. The equivalent extreme-value event and evaluation of the structural system reliability. *Structural Safety*, 29(2): 112–131, 2007.

Q. Li, C. Wang, and B. R. Ellingwood. Time-dependent reliability of aging structures in the presence of non-stationary loads and degradation. *Structural Safety*, 52:132–141, 2015b.

J. Luque and D. Straub. Reliability analysis and updating of deteriorating systems with dynamic Bayesian networks. *Structural Safety*, 62:34–46, 2016.

R. E. Melchers and A. T. Beck. *Structural Reliability Analysis and Prediction*. John Wiley & Sons, 2018.

A. Naess, B. Leira, and O. Batsevych. System reliability analysis by enhanced Monte Carlo simulation. *Structural Safety*, 31(5):349–355, 2009.

F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

V. Papadopoulos, D. G. Giovanis, N. D. Lagaros, and M. Papadrakakis. Accelerated subset simulation with neural networks for reliability analysis. *Computer Methods in Applied Mechanics and Engineering*, 223:70–80, 2012.

I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub. MCMC algorithms for subset simulation. *Probabilistic Engineering Mechanics*, 41:89–103, 2015.

M. Pellissetti, G. Schuëller, H. Pradlwarter, A. Calvi, S. Fransen, and M. Klein. Reliability analysis of spacecraft structures under static and dynamic loading. *Computers & Structures*, 84(21):1313–1325, 2006.

K.-K. Phoon. *Reliability-Based Design in Geotechnical Engineering: Computations and Applications*. CRC Press, 2008.

G. Pólya and G. Szegö. *Aufgaben und Lehrsätze aus der Analysis: Funktionentheorie· Nullstellen· Polynome· Determinanten· Zahlentheorie*, volume 74. Springer-Verlag, 2013.

J. Qin, K. Nishijima, and M. H. Faber. Extrapolation method for system reliability assessment: A new scheme. *Advances in Structural Engineering*, 15(11):1893–1909, 2012.

M. Romano, M. Losacco, C. Colombo, and P. Di Lizia. Impact probability computation of near-earth objects using Monte Carlo line sampling and subset simulation. *Celestial Mechanics and Dynamical Astronomy*, 132(8): 1–31, 2020.

R. Schneider, S. Thöns, and D. Straub. Reliability analysis and updating of deteriorating systems with subset simulation. *Structural Safety*, 64:20–36, 2017.

G. Schuëller, H. Pradlwarter, and P. Koutsourelakis. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19(4):463–474, 2004.

G. I. Schuëller and H. J. Pradlwarter. Benchmark study on reliability estimation in higher dimensions of structural systems–an overview. *Structural Safety*, 29(3):167–182, 2007.

J. Seplveda and M. H. Faber. Benchmark of emerging structural reliability methods. In *3th International Conference on Applications of Statistics and Probability in Civil Engineering*, 2019.

Y. Shi, Z. Lu, K. Zhang, and Y. Wei. Reliability analysis for structures with multiple temporal and spatial parameters based on the effective first-crossing point. *Journal of Mechanical Design*, 139(12), 2017.

J. Song and Z.-H. Wang. Evaluating the impact of built environment characteristics on urban boundary layer dynamics using an advanced stochastic approach. *Atmospheric Chemistry & Physics*, 16(10), 2016.

S. Song, Z. Lu, and H. Qiao. Subset simulation for structural reliability sensitivity analysis. *Reliability Engineering & System Safety*, 94(2):658–665, 2009.

M. G. Stewart and D. V. Rosowsky. Time-dependent reliability of deteriorating reinforced concrete bridge decks. *Structural Safety*, 20(1):91–109, 1998.

R. Stocki, K. Kolanek, J. Knabel, and P. Tauzowski. FE based structural reliability analysis using stand environment. *Computer Assisted Methods in Engineering and Science*, 16(1):35–58, 2017.

D. Straub. Stochastic modeling of deterioration processes through dynamic Bayesian networks. *Journal of Engineering Mechanics*, 135(10):1089–1099, 2009.

D. Straub and A. Der Kiureghian. Bayesian network enhanced with structural reliability methods: methodology. *Journal of engineering mechanics*, 136(10):1248–1258, 2010.

G. Su, L. Peng, and L. Hu. A gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis. *Structural Safety*, 68:97–109, 2017.

S. Sun and X. Li. Fast statistical analysis of rare circuit failure events via subset simulation in high-dimensional variation space. In *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 324–331. IEEE, 2014.

F. Tuyl, R. Gerlach, and K. Mengersen. A comparison of Bayes–Laplace, Jeffreys, and other priors: the case of zero events. *The American Statistician*, 62(1):40–44, 2008.

L. Tvedt. Distribution of quadratic forms in normal space application to structural reliability. *Journal of Engineering Mechanics*, 116(6):1183–1197, 1990.

C. Wang, L. Drees, and F. Holzapfel. Incident prediction using subset simulation. In *Proc. of ICAS 2014 29th Congress of the International Council of the Aeronautical Sciences*, pages 1–8, 2014.

P. Wei, Z. Lu, W. Hao, J. Feng, and B. Wang. Efficient sampling methods for global reliability sensitivity analysis. *Computer Physics Communications*, 183(8):1728–1743, 2012.

P. Wei, Z. Lu, and X. Yuan. Monte Carlo simulation for moment-independent sensitivity analysis. *Reliability Engineering & System Safety*, 110:60–67, 2013.

Q. Xiao. Evaluating correlation coefficient for Nataf transformation. *Probabilistic Engineering Mechanics*, 37:1–6, 2014.

S. Xiao, S. Oladyshkin, and W. Nowak. Reliability sensitivity analysis with subset simulation: application to a carbon dioxide storage problem. In *IOP Conference Series: Materials Science and Engineering*, volume 615, page 012051. IOP Publishing, 2019.

J. Yang and Z.-H. Wang. Physical parameterization and sensitivity of urban hydrological models: Application to green roof systems. *Building and Environment*, 75:250–263, 2014.

J. Yi, F. Wu, Q. Zhou, Y. Cheng, H. Ling, and J. Liu. An active-learning method based on multi-fidelity kriging model for structural reliability analysis. *Structural and Multidisciplinary Optimization*, pages 1–23, 2020.

W. Yun, Z. Lu, and X. Jiang. A modified importance sampling method for structural reliability and its global reliability sensitivity analysis. *Structural and Multidisciplinary Optimization*, 57(4):1625–1641, 2018a.

W. Yun, Z. Lu, Y. Zhang, and X. Jiang. An efficient global reliability sensitivity analysis algorithm based on classification of model output and subset simulation. *Structural Safety*, 74:49–57, 2018b.

D. Zhang, X. Han, C. Jiang, J. Liu, and Q. Li. Time-dependent reliability analysis through response surface method. *Journal of Mechanical Design*, 139(4), 2017.

Y.-G. Zhao and T. Ono. A general procedure for first/second-order reliabilitymethod (form/sorm). *Structural Safety*, 21(2):95–112, 1999.

Y.-G. Zhao and T. Ono. Moment methods for structural reliability. *Structural Safety*, 23(1):47–75, 2001.

E. Zio and N. Pedroni. Monte Carlo simulation-based sensitivity analysis of the model of a thermal–hydraulic passive system. *Reliability Engineering & System Safety*, 107:90–106, 2012.

K. M. Zuev and M. Beer. Reliability of critical infrastructure networks: Challenges. In *Resilience Engineering for Urban Tunnels*, pages 71–82. 2018.

K. M. Zuev, J. L. Beck, S.-K. Au, and L. S. Katafygiotis. Bayesian post-processor and other enhancements of subset simulation for estimating failure probabilities in high dimensions. *Computers & Structures*, 92:283–296, 2012.

K. M. Zuev, S. Wu, and J. L. Beck. General network reliability problem and its efficient solution by subset simulation. *Probabilistic Engineering Mechanics*, 40:25–35, 2015.

K. Zwirglmaier and D. Straub. A discretization procedure for rare events in Bayesian networks. *Reliability Engineering & System Safety*, 153:96–109, 2016.

Appendices

Appendix A

Necessity of the Independence Assumption

We provide a simple example of why we need the Independence Assumption (Assumption 2.2.1). Afterwards, we shortly discuss that also the Nataf transformation may not be helpful in such a model.

Example A.0.1 (Necessity of the Independence Assumption)**.** *In this example, we will consider discrete random variables for simplification, whereas a similar example can easily be constructed with continuous random variables. The model consists of a two dimensional random variable* $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ *with* $\mathbf{X}_1 \in \{-0.5, 0.5\}$ *and* $\mathbf{X}_2 \in \{0, 1\}$ *where we allow dependencies between* $\mathbf{X}_1$ *and* $\mathbf{X}_2$*, given by their conditional probabilities*

$$P(\mathbf{X}_1 = 0.5|\mathbf{X}_2 = 0) = P(\mathbf{X}_1 = -0.5|\mathbf{X}_2 = 1) = c$$

and

$$P(\mathbf{X}_1 = -0.5|\mathbf{X}_2 = 0) = P(\mathbf{X}_1 = 0.5|\mathbf{X}_2 = 1) = 1 - c$$

for some $c \in [0.1]$*. It is straightforward to check that this corresponds to the correlation* $\mathrm{Corr}(\mathbf{X}_1, \mathbf{X}_2) = 1 - 2c$ *between* $\mathbf{X}_1$ *and* $\mathbf{X}_2$*. The limit state function of the model is given by*

$$g : \mathbb{R}^2 \to \mathbb{R},\ g(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1 + \mathbf{x}_2$$

and the threshold is set to $b^* = 0$*. In this setting, we assume* $\mathbf{X}_2$ *as the dynamic variable* $\mathbf{X}_k = \mathbf{X}_2$*. We have to show that by calculating* $P(g(\mathbf{X}) < 0|\mathbf{X}_2 = y)$ *for all* $y \in D_2$ *with respect to a specific distribution of* $\mathbf{X}_2$ *does not*

generally allow to draw conclusions on the failure probability with respect to a different distribution of $\mathbf{X}_2$ *without additional evaluations of the limit state function* g*, when correlations are present. To do so, we introduce another constellation, where we assume a different distribution for* $\mathbf{X}_2$*, referring to the corresponding random variable as* $\bar{\mathbf{X}}_2$*. As a consequence, it should then be necessary to add more expensive reliability evaluations for estimation of the failure probability in the new setting. As distributions for the relevant model constellations, we consider*

$$P(\mathbf{X}_1 = -0.5) = P(\mathbf{X}_1 = 0.5) = 0.5$$

and either

$$P(\mathbf{X}_2 = 0) = P(\mathbf{X}_2 = 1) = 0.5$$

or

$$P(\bar{\mathbf{X}}_2 = -1) = P(\bar{\mathbf{X}}_2 = 0) = 0.5 \ .$$

For the newly introduced model, we again define conditional probabilities

$$P(\mathbf{X}_1 = 0.5|\bar{\mathbf{X}}_2 = -1) = P(\mathbf{X}_1 = -0.5|\bar{\mathbf{X}}_2 = 0) = c$$

and

$$P(\mathbf{X}_1 = 0.5|\bar{\mathbf{X}}_2 = 0) = P(\mathbf{X}_1 = -0.5|\bar{\mathbf{X}}_2 = -1) = 1 - c \ .$$

This setting again corresponds to the same correlation $\mathrm{Corr}(\mathbf{X}_1, \bar{\mathbf{X}}_2) = 1 - 2c$ *between* $\mathbf{X}_1$ *and the dynamic variable. When computing the conditional failure probabilities according to Equation 2.2 at zero, we then have*

$$P(\mathbf{X}_1 + \mathbf{X}_2 < 0|\mathbf{X}_2 = 0) = P(\mathbf{X}_1 = -0.5|\mathbf{X}_2 = 0) = c$$

and

$$P(\mathbf{X}_1 + \bar{\mathbf{X}}_2 < 0|\bar{\mathbf{X}}_2 = 0) = P(\mathbf{X}_1 = -0.5|\bar{\mathbf{X}}_2 = 0) = 1 - c$$

which only coincide if $c = 0.5$*. Since this corresponds to zero correlation, the model only allows to draw conclusions for uncorrelated random variables at this point.*

For completeness, we also shortly discuss the effect of the Nataf transformation in such a model in a degenerate setting, representing the discrete random variables in the original problem formulation as transformation of independent standard normally distributed random variables and modifying

the limit state equation accordingly. Following Example A.0.1, we get (compare Xiao (2014)) $\mathbf{x}_1 = F_1^{-1}(\phi(\mathbf{u}_1))$ and $\mathbf{x}_2 = F_2^{-1}\left(\phi\left(\rho_Z \mathbf{u}_1 + \sqrt{1-\rho_Z^2}\mathbf{u}_2\right)\right)$ or equivalently a reformulated limit state equation in standard normal space (U-space), here given by

$$g_U(\mathbf{u}_1, \mathbf{u}_2) = F_1^{-1}(\phi(\mathbf{u}_1)) + F_2^{-1}\left(\phi\left(\rho_Z \mathbf{u}_1 + \sqrt{1-\rho_Z^2}\mathbf{u}_2\right)\right) \ . \qquad \text{(A.1)}$$

The random variables $\mathbf{U}_1$, $\mathbf{U}_2$ in U-space are independent and standard normally distributed. Here ϕ represents the cumulative distribution function (cdf) of the standard normal distribution and F_1^{-1} and F_2^{-1} the inverse cdfs of $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. We do not further specify the correlation $\rho_Z \in [0,1]$ between $\mathbf{Z}_1$ and $\mathbf{Z}_2$, the random variables in Z-space (standard normally but potentially correlated random variables), since we only need to consider the difference between the cases $\rho_z < 0$, $\rho_z = 0$ and $\rho_z > 0$. The limit state function in the second constellation is derived accordingly and given by replacing the inverse cdf of $\mathbf{X}_2$ by the corresponding inverse cdf $\bar{F}_2^{-1}$ of $\bar{\mathbf{X}}_2$ in $g_U(\mathbf{u}_1, \mathbf{u}_2)$. For $\mathbf{z}_1 = \mathbf{u}_1$ and $\mathbf{z}_2 = \rho_Z \mathbf{u}_1 + \sqrt{1-\rho_Z^2}\mathbf{u}_2$, the inverse cdfs are given by

- $F_1^{-1}(\phi(\mathbf{z}_1)) = -0.5$ for $\phi(\mathbf{z}_1) \in [0, 0.5]$ and $F_1^{-1}(\phi(\mathbf{z}_1)) = 0.5$ for $\phi(\mathbf{z}_1) \in (0.5, 1]$,
- $F_2^{-1}(\phi(\mathbf{z}_2)) = 0$ for $\phi(\mathbf{z}_2) \in [0, 0.5]$ and $F_2^{-1}(\phi(\mathbf{z}_2)) = 1$ for $\phi(\mathbf{z}_2) \in (0.5, 1]$,
- $\bar{F}_2^{-1}(\phi(\mathbf{z}_2)) = -1$ for $\phi(\mathbf{z}_2) \in [0, 0.5]$ and $\bar{F}_2^{-1}(\phi(\mathbf{z}_2)) = 0$ for $\phi(\mathbf{z}_2) \in (0.5, 1]$.

So, again we fix the dynamic variable and try to use results of one model for the other model. However, the limit states differ substantially under dependencies, as the inverse cdfs differ. To state out this point more clearly, consider that $F_2^{-1}(\phi(\mathbf{z}_2)) = 0$ for $\phi(\mathbf{z}_2) \in [0, 0.5]$ and $\bar{F}_2^{-1}(\phi(\mathbf{z}_2)) = 0$ for $\phi(\mathbf{z}_2) \in (0.5, 1]$ which means the second summand of the limit states takes the same value in both models only if we have $\phi\left(\rho_Z \mathbf{u}_1 + \sqrt{1-\rho_Z^2}\mathbf{u}_2\right) \in [0, 0.5]$ in the first model and $\phi\left(\rho_Z \mathbf{u}_1 + \sqrt{1-\rho_Z^2}\mathbf{u}_2\right) \in (0.5, 1]$ in the second one. It is easy to check that, in the first model this corresponds to the three cases

- $\mathbf{u}_1 \leq \frac{-\sqrt{1-\rho_Z^2}}{\rho_Z}\mathbf{u}_2$ with $\frac{-\sqrt{1-\rho_Z^2}}{\rho_Z} < 0$, if $\rho_z < 0$

- $\mathbf{u}_2 \leq 0$, if $\rho_z = 0$
- $\mathbf{u}_1 \leq \frac{-\sqrt{1-\rho_Z^2}}{\rho_Z}\mathbf{u}_2$ with $\frac{-\sqrt{1-\rho_Z^2}}{\rho_Z} > 0$, if $\rho_z > 0$

where in the second model we have to alter the inequality by replacing '$\leq$' with '$>$'.Thus, when values of the second summand of the limit state equations coincide for some $\mathbf{u}_2$, the realizations of $\mathbf{u}_1$ must differ, possibly leading to a different outcome of the limit state function value. If no correlations are present (case $\rho_z = 0$), then $\mathbf{u}_1$ does not contribute to the outcome of the second summand so that we may have same $\mathbf{u}_1$ for the same values of the second summand in both models. In general, it is thus not clear how evaluating one of the limit state equations could provide certain information on the outcome of the other one. Additional, correlations might exist between many random variables and when assuming a high complexity of the limit state function, substantial changes might occur in the limit state function when switching probability distributions of the dynamic variable. Thus, re-evaluation would become necessary also when using a Nataf transformation. In case of the Nataf transformation, dependencies are assigned to the limit state function, but are still present.

In summary, we can end up with different conditional probabilities in Equation 2.2, for different dependency structures. When considering (2.2) for several f_k, it is therefore useful to impose Assumption 2.2.1 so that one can separate the problem.

Appendix B

Properties of the Beta Function and Distribution

Definition B.0.1. *The Beta function is given by*

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

for complex numbers α, β with positive real parts and $\Gamma(\cdot)$ the Gamma function.

Lemma B.0.2. *For real numbers $\alpha > 0$, $\beta > 0$, we have*

$$B(\alpha + 1, \beta) = B(\alpha, \beta) \cdot \frac{\alpha}{\alpha + \beta}$$

and

$$B(\alpha, \beta + 1) = B(\alpha, \beta) \cdot \frac{\alpha}{\alpha + \beta} \ .$$

Proof. The Gamma function satisfies the recurrence relation $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ (see e.g. Olver et al. (2010)). Thus, we have

$$B(\alpha + 1, \beta) = \frac{\Gamma(\alpha + 1)\Gamma(\beta)}{\Gamma(\alpha + \beta + 1)} = \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha + \beta)\Gamma(\alpha + \beta)} = B(\alpha, \beta) \cdot \frac{\alpha}{\alpha + \beta} \ ,$$

which yields the first conclusion. The second equation is shown alike. □

Lemma B.0.3. *[Expected Value of an Inverted Beta variable] For a Beta distributed random variable $p \sim Beta(\alpha, \beta)$, we have*

$$E\left[\frac{1}{p}\right] = \frac{\alpha + \beta - 1}{\alpha - 1}$$

a.s. $(p \neq 0)$.

Proof. First, by definition of the expectation, we have

$$E\left[\frac{1}{p}\right] \overset{a.s.}{=} \int_{0<p\leq 1} \frac{1}{p} P(p) dp \ .$$

Plugging in the pdf (Definition 4.2.10), with B the Beta function, gives

$$\begin{aligned} E\left[\frac{1}{p}\right] &\overset{a.s.}{=} \int_{0<p\leq 1} \frac{1}{p} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} dp \ . \\ &= \int_{0<p\leq 1} \frac{p^{\alpha-2}(1-p)^{\beta-1}}{B(\alpha,\beta)} dp \\ &= \frac{B(\alpha-1,\beta)}{B(\alpha,\beta)} \int_{0<p\leq 1} \frac{p^{\alpha-2}(1-p)^{\beta-1}}{B(\alpha-1,\beta)} dp \ . \end{aligned}$$

Thus, as a pdf is integrated, we have

$$E\left[\frac{1}{p}\right] \overset{a.s.}{=} \frac{B(\alpha-1,\beta)}{B(\alpha,\beta)} \ .$$

Using $B(\alpha,\beta) = B(\alpha-1,\beta)\frac{\alpha-1}{\alpha-1+\beta}$ (see Lemma B.0.2) yields the result. □

Appendix C

Additional Proofs for Chapter 5

Proof. (**Lemma 5.3.30**): To determine the difference between real derivative and tangent approximation, we use the following Taylor approximations:

$$q_{w+1} = q_w + hq_w^{(1)} + \frac{h^2}{2}q_w^{(2)} + \frac{h^3}{3!}q_w^{(3)} + \frac{h^4}{4!}q_w^{(4)} + \frac{h^5}{5!}q^{(5)}(\tau_1) \tag{C.1}$$

for some $\tau_1 \in [y_w, y_{w+1}]$ and

$$q_{w-1} = q_w - hq_w^{(1)} + \frac{h^2}{2}q_w^{(2)} - \frac{h^3}{3!}q_w^{(3)} + \frac{h^4}{4!}q_w^{(4)} - \frac{h^5}{5!}q^{(5)}(\tau_2) \tag{C.2}$$

for some $\tau_2 \in [y_{w-1}, y_w]$. Then subtracting (C.2) from (C.1) yields

$$q_{w+1} - q_{w-1} = 2hq_w^{(1)} + \frac{2h^3}{3!}q_w^{(3)} + \frac{h^5}{5!}\left(q^{(5)}(\tau_1) + q^{(5)}(\tau_2)\right) \ .$$

Reordering yields

$$\frac{q_{w+1} - q_{w-1}}{2h} = q_w^{(1)} + \frac{h^2}{6}q_w^{(3)} + \frac{h^4}{240}\left(q^{(5)}(\tau_1) + q^{(5)}(\tau_2)\right) \tag{C.3}$$

where the left hand side is just m_w of (3FD). If instead using Taylor formula up to terms of third order, we can similarly derive

$$\frac{q_{w+1} - q_{w-1}}{2h} = q_w^{(1)} + \frac{h^2}{6}\left(q^{(3)}(\tau_1^*) + q^{(3)}(\tau_2^*)\right)$$

for some $\tau_1^* \in [y_w, y_{w+1}]$ and $\tau_2^* \in [y_{w-1}, y_w]$. We thus also have by continuity

$$\frac{q_{w+1} - q_{w-1}}{2h} = q_w^{(1)} + \frac{h^2}{3} q^{(3)}(\tau_{w,3FD})$$

for an $\tau_{w,3FD} \in [y_{w-1}, y_{w+1}]$. This proofs the first part (3FD). For (5FD) we additionally need

$$q_{w+2} = q_w + 2h q_w^{(1)} + \frac{4h^2}{2} q_w^{(2)} + \frac{8h^3}{3!} q_w^{(3)} + \frac{16h^4}{4!} q_w^{(4)} + \frac{32h^5}{5!} q^{(5)}(\tau_3) \quad \text{(C.4)}$$

and

$$q_{w-2} = q_w - 2h q_w^{(1)} + \frac{4h^2}{2} q_w^{(2)} - \frac{8h^3}{3!} q_w^{(3)} + \frac{16h^4}{4!} q_w^{(4)} - \frac{32h^5}{5!} q^{(5)}(\tau_4) \quad \text{(C.5)}$$

for some $\tau_3 \in [y_w, y_{w+2}]$, $\tau_4 \in [y_{w-2}, y_w]$, where we used $(y_w - y_{w-2})^i = (2h)^i = 2^i h^i$. Now we follow the same procedure as above again and subtract (C.5) from (C.4) so that

$$q_{w+2} - q_{w-2} = 4h q_w^{(1)} + \frac{16h^3}{3!} q_w^{(3)} + \frac{32h^5}{5!} \left(q^{(5)}(\tau_3) + q^{(5)}(\tau_4) \right) .$$

Reformulation yields

$$\frac{q_{w+2} - q_{w-2}}{4h} = q_w^{(1)} + \frac{2h^2}{3} q_w^{(3)} + \frac{h^4}{15} \left(q^{(5)}(\tau_3) + q^{(5)}(\tau_4) \right) . \quad \text{(C.6)}$$

Now $4 \cdot (\text{C.3}) - 1 \cdot (\text{C.6})$ eliminates the terms of order h^2

$$\begin{aligned} \frac{q_{w-2} - 8q_{w-1} + 8q_{w+1} - q_{w+2}}{4h} =& 3q_w^{(1)} + \frac{h^4}{60} \Big(\left(q^{(5)}(\tau_1) + q^{(5)}(\tau_2) \right) \\ & - 4 \left(q^{(5)}(\tau_3) + q^{(5)}(\tau_4) \right) \Big) . \end{aligned}$$

Division by 3 on both sides then yields the claim for the (5FD) approximation

$$\begin{aligned} \frac{q_{w-2} - 8q_{w-1} + 8q_{w+1} - q_{w+2}}{12h} =& q_w^{(1)} + \frac{h^4}{180} \Big(\left(q^{(5)}(\tau_1) + q^{(5)}(\tau_2) \right) \\ & - 4 \left(q^{(5)}(\tau_3) + q^{(5)}(\tau_4) \right) \Big) . \end{aligned}$$

This yields the first part for 5FD. By continuity, we then find a $\tau_{1,2} \in [y_{w-1}, y_{w+1}]$ and $\tau_{3,4} \in [y_{w-2}, y_{w+2}]$ so that

$$\left(q^{(5)}(\tau_1) + q^{(5)}(\tau_2) \right) - 4 \left(q^{(5)}(\tau_3) + q^{(5)}(\tau_4) \right) = q^{(5)}(\tau_{1,2}) - 4q^{(5)}(\tau_{3,4}) .$$

Now, accounting for the four different cases of the signs of these higher order derivatives, we can bound the absolute value of the term under inclusion of all the cases by

$$|q^{(5)}(\tau_{1,2}) - 4q^{(5)}(\tau_{3,4})| \in [0, |q^{(5)}(\tau_{1,2})| + 4|q^{(5)}(\tau_{3,4})|] \\ \subseteq [0, 5\max(|q^{(5)}(\tau_{1,2})|, |q^{(5)}(\tau_{3,4})|)] \ .$$

Defining $\tau_{w,5FD} := argmax(|q^{(5)}(\tau_{1,2})|, |q^{(5)}(\tau_{3,4})|)$ then yields the second part for 5FD

$$\frac{q_{w-2} - 8q_{w-1} + 8q_{w+1} - q_{w+2}}{12h} \in \left[q_w^{(1)} - \left|\frac{h^4}{36}q^{(5)}(\tau_{w,5FD})\right|, q_w^{(1)} + \left|\frac{h^4}{36}q^{(5)}(\tau_{w,5FD})\right|\right]$$

for a $\tau_{w,5FD} \in [y_{w-2}, y_{w+2}]$. □

Proof. (**Lemma 5.3.31**):

$$m_w = \frac{2d_{w-1}d_w}{d_{w-1} + d_w}$$

Again by Taylor approximation, we have

$$q_{w+1} = q_w + hq_w^{(1)} + \frac{h^2}{2}q_w^{(2)} + \frac{h^3}{3!}q^{(3)}(\tau_3)$$

for some $\tau_3 \in [y_w, y_{w+1}]$ and

$$q_{w-1} = q_w - hq_w^{(1)} + \frac{h^2}{2}q_w^{(2)} - \frac{h^3}{3!}q^{(3)}(\tau_4)$$

for some $\tau_4 \in [y_{w-1}, y_w]$. Plugging into the two point finite differences d_{w-1}, d_w results in

$$d_{w-1} = q_w^{(1)} - \frac{h}{2}q_w^{(2)} + \frac{h^2}{6}q^{(3)}(\tau_4)$$

$$d_w = q_w^{(1)} + \frac{h}{2}q_w^{(2)} + \frac{h^2}{6}q^{(3)}(\tau_3) \ .$$

We thus have

$$\frac{2d_{w-1}d_w}{d_{w-1}+d_w} = \left(2(q_w^{(1)})^2 + \frac{h^2}{3}q_w^{(1)}\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right) - \frac{h^2}{2}(q_w^{(2)})^2 + \frac{h^3}{6}q_w^{(2)}\left(q^{(3)}(\tau_4)-q^{(3)}(\tau_3)\right) + \frac{h^4}{18}q^{(3)}(\tau_3)q^{(3)}(\tau_4)\right) \cdot \frac{1}{2q_w^{(1)}+\frac{h^2}{6}\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right)}.$$

Rewriting yields

$$\left(q_w^{(1)}\left(12q_w^{(1)}+h^2\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right)\right) + h^2q_w^{(1)}\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right) - 3h^2(q_w^{(2)})^2 + h^3q_w^{(2)}\left(q^{(3)}(\tau_4)-q^{(3)}(\tau_3)\right) + \frac{h^4}{3}q^{(3)}(\tau_3)q^{(3)}(\tau_4)\right) \cdot \frac{1}{12q_w^{(1)}+h^2\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right)}.$$

Next, we may simplify to receive

$$q_w^{(1)} + \left(h^2q_w^{(1)}\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right) - 3h^2(q_w^{(2)})^2 + h^3q_w^{(2)}\left(q^{(3)}(\tau_4)-q^{(3)}(\tau_3)\right) + \frac{h^4}{3}q^{(3)}(\tau_3)q^{(3)}(\tau_4)\right) \cdot \frac{1}{12q_w^{(1)}+h^2\left(q^{(3)}(\tau_3)+q^{(3)}(\tau_4)\right)}$$

which we may rewrite again to get the result. □

Proof. (**Proposition 5.3.34**): By definition of HM (5.16)

$$m_w = \frac{2d_{w-1}d_w}{d_{w-1}+d_w}$$

and the two point finite differences, the estimation error is given by

$$\begin{aligned}\hat{m}_w - m_w =& \frac{2(\hat{q}_w\hat{q}_{w+1} - \hat{q}_{w-1}\hat{q}_{w+1} - (\hat{q}_w)^2 + \hat{q}_{w-1}\hat{q}_w)}{h(\hat{q}_{w+1} - \hat{q}_{w-1})} \\ & - \frac{2(q_w q_{w+1} - q_{w-1}q_{w+1} - (q_w)^2 + q_{w-1}q_w)}{h(q_{w+1} - q_{w-1})} \ .\end{aligned}$$

To proceed further, it is unfortunately not possible to approximate the denominator $h(\hat{q}_{w+1}-\hat{q}_{w-1}) = h(q_{w+1}-q_{w-1})+h(\mathcal{E}_{w+1}-\mathcal{E}_{w-1})$ by $h(d_{w+1}-d_{w-1})$, because the error of this simplification would be $O(1/\sqrt{N})$ and can therefore not be neglected. An explicit result by straightforward calculation is given as

$$\begin{aligned}&\frac{2}{h(q_{w-1} - q_{w+1})^2 + h(q_{w-1} - q_{w+1})(\mathcal{E}_{w-1} - \mathcal{E}_{w+1})}\cdot \\ &\Big(\mathcal{E}_{w-1}(-q_w^2 - q_{w+1}^2 + 2q_w q_{w+1}) \\ &+\mathcal{E}_w(-q_{w-1}^2 + q_{w+1}^2 + 2q_{w-1}q_w - 2q_w q_{w+1}) \\ &+\mathcal{E}_{w+1}(q_{w-1}^2 + q_w^2 - 2q_{w-1}q_w)\Big)\end{aligned}$$

Now the nominator is $O(1/\sqrt{N})$ as a linear combination of $O(1/\sqrt{N})$ Gaussians. The denominator $h(q_{w-1} - q_{w+1})^2 + h(q_{w-1} - q_{w+1})(\mathcal{E}_{w-1} - \mathcal{E}_{w+1})$ on the other hand is given by the sum of the constant $h(q_{w-1} - q_{w+1})^2$ and some $O(1/\sqrt{N})$ term. Thus, for high N, we may neglect the stochastic term in the denominator so that it suffices to consider

$$\begin{aligned}&\frac{2}{h(q_{w-1} - q_{w+1})^2}\cdot \\ &\Big(\mathcal{E}_{w-1}(-q_w^2 - q_{w+1}^2 + 2q_w q_{w+1}) \\ &+\mathcal{E}_w(-q_{w-1}^2 + q_{w+1}^2 + 2q_{w-1}q_w - 2q_w q_{w+1}) \\ &+\mathcal{E}_{w+1}(q_{w-1}^2 + q_w^2 - 2q_{w-1}q_w)\Big)\end{aligned}$$

for derivation of asymptotic results.

Now, by homogeneous subsets assumption (Assumption 5.3.5) we can further simplify, using $q_{w-1} = q_w p_0 = q_{w+1}p_0^2$ to get

$$\begin{aligned}&\frac{2}{h(1 - 2p_0^2 + p_0^4)}\cdot\Big(\mathcal{E}_{w-1}(-p_0^2 + 2p_0^3 - p_0^4) + \mathcal{E}_w(-1 + 2p_0 - 2p_0^3 + p_0^4) \\ &\qquad + \mathcal{E}_{w+1}(1 - 2p_0 + p_0^2)\Big) \ .\end{aligned}$$

This again can be simplified to

$$\frac{2\left(-p_0^2\mathcal{E}_{w-1}+\mathcal{E}_w(p_0^2-1)+\mathcal{E}_{w+1}\right)}{h(p_0+1)^2}=\frac{2\left(-p_0^2\mathcal{E}_{w-1}+\mathcal{E}_w(p_0^2-1)+\mathcal{E}_{w+1}\right)}{h(1+2p_0+p_0^2)}\ .$$

□

Relation of the Variances in 3FD and HM, Remark 5.3.35 First, we only look at the numerator of $\hat{m}_{w,err}^{HM}$

$$\frac{2\left(-p_0^2\mathcal{E}_{w-1}+\mathcal{E}_w(p_0^2-1)+\mathcal{E}_{w+1}\right)}{h(1+2p_0+p_0^2)}$$

which then results, under replacement of the estimation errors by (5.20), in

$$\begin{aligned}2\left(-\sum_{i=1}^{w-1}\sigma_i p_0^w Z_i-\sum_{i=1}^{w}\sigma_i p_0^{w-1}Z_i+\sum_{i=1}^{w}\sigma_i p_0^{w+1}Z_i+\sum_{i=1}^{w+1}\sigma_i p_0^w Z_i\right)\\ =2\left(\sum_{i=1}^{w}\sigma_i Z_i\cdot(p_0^{w+1}-p_0^{w-1})+\sum_{i=w}^{w+1}\sigma_i Z_i p_0^w\right)\ .\end{aligned}$$

Similarly, for the numerator of $\hat{m}_{w,err}^{3FD}$, we get

$$\begin{aligned}\mathcal{E}_{w+1}-\mathcal{E}_{w-1}&=\sum_{i=1}^{w+1}\sigma_i p_0^w Z_i-\sum_{i=1}^{w-1}\sigma_i p_0^{w-2}Z_i\\ &=\left(\sum_{i=1}^{w-1}\sigma_i Z_i\right)p_0^{w-2}\left(p_0^2-1\right)+\sum_{i=w}^{w+1}\sigma_i Z_i p_0^w\ .\end{aligned}$$

We can then compare the variances of the numerators of $\hat{m}_{w,err}^{HM}$ and $\hat{m}_{w,err}^{3FD}$

$$\begin{aligned}&\frac{\mathrm{Var}\left(2\left(-p_0^2\mathcal{E}_{w-1}+\mathcal{E}_w(p_0^2-1)+\mathcal{E}_{w+1}\right)\right)}{\mathrm{Var}\left(\mathcal{E}_{w+1}-\mathcal{E}_{w-1}\right)}\\ &=\frac{4\left(w\sigma^2(p_0^{w+1}-p_0^{w-1})^2+2\sigma^2p_0^{2w}\right)}{(w-1)\sigma^2p_0^{2(w-2)}(p_0^2-1)^2+2\sigma^2p_0^{2w}}\\ &=\frac{4p_0^2p_0^{2w-4}\left(wp_0^4-2wp_0^2+w+2p_0^2\right)}{p_0^{2w-4}\left(-1+w+2p_0^2-2wp_0^2+p_0^4+wp_0^4\right)}\\ &=\frac{4p_0^2\left(wp_0^4-2wp_0^2+w+2p_0^2\right)}{-1+w+2p_0^2-2wp_0^2+p_0^4+wp_0^4}\end{aligned}$$

where we used the fact that a sum of independent Gaussians yields a Gaussian distributed random variable again with variance given by the sum of the variances of the individual Gaussian variables. The mean value of all $Z_1, ..., Z_{w+1}$ is zero so that the resulting Gaussians also have mean value zero. Note that the p_0^2 term in the numerator indicates small values of the fraction for small p_0. Also adding the denominators to the analysis yields

$$\frac{\mathrm{Var}(\hat{m}_{w,err}^{HM})}{\mathrm{Var}(\hat{m}_{w,err}^{3FD})} = \frac{4p_0^2\left(wp_0^4 - 2wp_0^2 + w + 2p_0^2\right)}{-1 + w + 2p_0^2 - 2wp_0^2 + p_0^4 + wp_0^4} \cdot \frac{4}{(p_0+1)^4} \tag{C.7}$$

The resulting term in (C.7) is not straightforward to interpret, but analysis only needs to cover specific cases so that it is sufficient to numerically derive a conclusion here.

Additional Information to the Proof of Proposition 5.3.37 In the following, we add more details to the derivation of (5.23) and (5.22) given in Proposition 5.3.37. First, we have to proceed as given in the following example, using Proposition 5.3.6.

$$\begin{aligned}
E\left[e_{w,4}^2\right] &= E\left[\left(\frac{1}{2}(\mathcal{E}_{w+2} - \mathcal{E}_w)\right)^2\right] \\
&= \frac{1}{4}\left(E\left[\mathcal{E}_{w+2}^2 - 2\mathcal{E}_{w+2}\mathcal{E}_w + \mathcal{E}_w^2\right]\right) \\
&= \frac{1}{4}\left(E\left[\mathcal{E}_{w+2}^2\right] - 2E\left[\mathcal{E}_{w+2}\mathcal{E}_w\right] + E\left[\mathcal{E}_w^2\right]\right) \\
&= \mathrm{Var}(\mathcal{E}_{k+2}) - 2\mathrm{Cov}(\mathcal{E}_{k+2}, \mathcal{E}_k) + \mathrm{Var}(\mathcal{E}_k) \\
&= \mathrm{Var}(\mathcal{E}_{k+2}) - 2\mathrm{Var}(\mathcal{E}_{k+2})p_0^2 + \mathrm{Var}(\mathcal{E}_k)
\end{aligned}$$

So, we use linearity of the expected value and then evaluate all individual terms as in the given example. We then have

$$\begin{aligned}
h \cdot E\Bigg[&\frac{2e_{w,4}^2 + (3e_{w,3} - 22e_{w,2} - 13e_{w,1})e_{w,4} + 2e_{w,3}^2}{210} \\
&\frac{-(13e_{w,2} + 22e_{w,1})e_{w,3} + 78e_{w,2}^2 + 54e_{w,1}e_{w,2} + 78e_{w,1}^2}{210}\Bigg] \\
= h\frac{1}{210}\cdot\Bigg(&\frac{1}{2}\left(\mathrm{Var}(\mathcal{E}_{w+2}) - 2\mathrm{Var}(\mathcal{E}_w)p_0^2 + \mathrm{Var}(\mathcal{E}_w)\right) \\
&-\frac{3}{4}\left(\mathrm{Var}(\mathcal{E}_{w+1})p_0 - \mathrm{Var}(\mathcal{E}_w)p_0 - \mathrm{Var}(\mathcal{E}_{w-1})p_0^3 + \mathrm{Var}(\mathcal{E}_{w-1})p_0\right) \\
&-11\left(\mathrm{Var}(\mathcal{E}_{w+1})p_0 - \mathrm{Var}(\mathcal{E}_w)p_0\right) - \frac{13}{2}\left(\mathrm{Var}(\mathcal{E}_w)p_0^2 - \mathrm{Var}(\mathcal{E}_w)\right) \\
&+\frac{1}{2}\left(\mathrm{Var}(\mathcal{E}_{w+1}) - 2\mathrm{Var}(\mathcal{E}_{w-1})p_0^2 + \mathrm{Var}(\mathcal{E}_{w-1})\right) \\
&+\frac{13}{2}\left(\mathrm{Var}(\mathcal{E}_{w+1}) - \mathrm{Var}(\mathcal{E}_{w-1})p_0^2\right) + 11\left(\mathrm{Var}(\mathcal{E}_w)p_0 - \mathrm{Var}(\mathcal{E}_{w-1})p_0\right) \\
&+78\mathrm{Var}(\mathcal{E}_{w+1}) + 54\mathrm{Var}(\mathcal{E}_w)p_0 + 78\mathrm{Var}(\mathcal{E}_w)\Bigg)\,.
\end{aligned}$$

Next, we also know $\mathrm{Var}(\mathcal{E}_w) = w\sigma^2 p_0^{2(w-1)}$ by Proposition 5.3.6 so that

$$\begin{aligned}
\mathrm{Var}(\mathcal{E}_{w-1}) &= \frac{w-1}{w}p_0^{-2}\mathrm{Var}(\mathcal{E}_w) \\
\mathrm{Var}(\mathcal{E}_{w+1}) &= \frac{w+1}{w}p_0^{2}\mathrm{Var}(\mathcal{E}_w) \\
\mathrm{Var}(\mathcal{E}_{w+2}) &= \frac{w+2}{w}p_0^{4}\mathrm{Var}(\mathcal{E}_w)
\end{aligned}$$

hold. This allows to deduce

$$\begin{aligned}
h\frac{\mathrm{Var}(\mathcal{E}_w)}{210}\cdot\Bigg(&\frac{1}{2}\frac{w+2}{w}p_0^4 - p_0^2 + \frac{1}{2} - \frac{3}{4}p_0^3\frac{w+1}{w} + \frac{3}{4}p_0 + \frac{3}{4}p_0\frac{w-1}{w} \\
&-\frac{3}{4}p_0^{-1}\frac{w-1}{w} - 11\frac{w+1}{w}p_0^3 + 11p_0 - \frac{13}{2}p_0^2 + \frac{13}{2} \\
&+\frac{1}{2}\frac{w+1}{w}p_0^2 - \frac{w-1}{w} + \frac{1}{2}\frac{w-1}{w}p_0^{-2}\frac{13}{2}\frac{w+1}{w}p_0^2 - \frac{13}{2}\frac{w-1}{w} + 11p_0 \\
&-11\frac{w-1}{w}p_0^{-1} + 78\frac{w+1}{w}p_0^2 + 54p_0 + 78\Bigg)\,.
\end{aligned}$$

Reordering yields

$$h\frac{\mathrm{Var}(\mathcal{E}_w)}{210}\cdot\Bigg(\frac{1}{2}\frac{w+2}{w}p_0^4-\frac{47}{4}\frac{w+1}{w}p_0^3+85\frac{w+1}{w}p_0^2-\frac{15}{2}p_0^2+\frac{3}{4}\frac{w-1}{w}p_0 \\ +76p_0-\frac{47}{4}\frac{w-1}{w}p_0^{-1}+\frac{1}{2}\frac{w-1}{w}p_0^{-2}-\frac{15}{2}\frac{w-1}{w}+85\Bigg)\,.$$

Scientific and Professional Career

Scientific Career

Since 08/2017	Ph.D. student of Prof. Dr. Jörn Sass, in the research training group RTG 1932 'Stochastic Models for Innovations in the Engineering Sciences', Technische Universität Kaiserslautern, Germany
10/2014-06/2017	Master studies in Business Mathematics with specialization Financial Mathematics (degree: M. Sc.), Technische Universität Kaiserslautern, Germany
04/2011-10/2014	Bachelor studies in Mathematics with minor Economics (degree: B. Sc.), Technische Universität Kaiserslautern, Germany

Professional Career

Since 10/2020	Founder and IT-Consultant, inTvestment UG, Germany
04/2017-07/2017	Trainee in Business-/Data-Analytics, 1&1 Telecommunication SE, Germany
05/2016-09/2016	Trainee in Risk Management & Control, Assenagon Asset Management S.A., Luxemburg

Wissenschaftlicher und Beruflicher Werdegang

Wissenschaftlicher Werdegang

Seit 08/2017	Doktorand von Prof. Dr. Jörn Sass, im Graduiertenkolleg GRK 1932 'Stochastic Models for Innovations in the Engineering Sciences', Technische Universität Kaiserslautern, Deutschland
10/2014-06/2017	Masterstudium in Wirtschaftsmathematik mit Vertiefung Finanzmathematik (Abschluss: M. Sc.), Technische Universität Kaiserslautern, Deutschland
04/2011-10/2014	Bachelorstudium in Mathematik mit Nebenfach Wirtschaftswissenschaften (Abschluss: B. Sc.), Technische Universität Kaiserslautern, Deutschland

Beruflicher Werdegang

Seit 10/2020	Gründer und IT-Consultant, inTvestment UG, Deutschland
04/2017-07/2017	Werkstudent in Business-/Data-Analytics, 1&1 Telecommunication SE, Deutschland
05/2016-09/2016	Praktikant in Risk Management & Control, Assenagon Asset Management S.A., Luxemburg