Jeffreys' prior is the Hausdorff measure for the Hellinger and Kullback-Leibler distances

Werner Doster
Fachbereich Mathematik, Universität Kaiserslautern
67663 Kaiserslautern, Germany
e-mail: doster@btbnet.de
September 24, 1998

Abstract

On a family $\mathcal{F} := \left\{ P_\vartheta \mid \vartheta \in \Theta \subset I\!R^k \right\}$ of probability measures on a measure space $(\Omega, \mathcal{A})$ we consider the Hellinger and Kullback-Leibler distances. We show that under suitable regularity conditions Jeffreys' prior is proportional to the k-dimensional Hausdorff measure w.r.t. Hellinger distance respectively to the $\frac{k}{2}$-dimensional Hausdorff measure w.r.t. Kullback-Leibler distance. The proof is based on an area-formula for the Hausdorff measure w.r.t. to generalized distances.

Keywords: Hausdorff measure, Jeffreys' prior, area-formula, Hellinger distance, Kullback-Leibler distance, Fisher Information
AMS classification numbers: 28A78,

Jeffreys' prior is the Hausdorff measure for the Hellinger and Kullback-Leibler distances

Werner Doster
Fachbereich Mathematik, Universität Kaiserslautern
67663 Kaiserslautern, Germany
e-mail: doster@btbnet.de
September 24, 1998

1 Indroduction

We consider a statistical experiment given by a family $\mathcal{F} := \{P_\vartheta \mid \vartheta \in \Theta \subset \mathbb{R}^k\}$ of probability measures on a measure space $(\Omega, \mathcal{A})$ dominated by a σ-finite measure μ and we assume $P_\vartheta \neq P_{\vartheta'}$ if $\vartheta \neq \vartheta'$. By $\{f_\vartheta \mid \vartheta \in \Theta \subset \mathbb{R}^k\} \subset L^1(\mu)$ we denote the corresponding family of densities. On $\mathcal{F}$ there are several ways to measure the distance between two probability measures. In this paper, we consider the Hellinger metric (cf. [5, I.7.6]) $d(\vartheta, \vartheta') := \sqrt{\int_\Omega \left(\sqrt{f_\vartheta} - \sqrt{f_{\vartheta'}}\right)^2 d\mu}$ and the Kullback-Leibler distance (cf. [9]) $\mathcal{K}(\vartheta, \vartheta') := \int_\Omega f_\vartheta \ln \frac{f_\vartheta}{f_{\vartheta'}} d\mu$.
The concept of Hausdorff measure gives a natural Borel regular outer measure on a metric space (cf. [12]). In our context this could be interpreted as a natural prior distribution on the parameter set Θ. In [8, Theorem 7] and independently in [3, Theorem 2.3.2] an area-formula for Banach space valued functions is proved. So, if we assume $\Phi : \Theta \longrightarrow L^2(\mu),\ \vartheta \mapsto \sqrt{f_\vartheta}$ to be locally Lipschitz-continuous the area-formula yields that k-dimensional Hausdorff measure w.r.t Hellinger distance is proportional to Jeffreys' prior (cf. [6, Section 3.10]). To prove the analogous result for Kullback-Leibler distance we first introduce the concept of Hausdorff measure w.r.t. generalized distances. Similar to the proof of the area-formula in [8] we get an area-formula for the Hausdorff measure w.r.t. such distances. Under suitable regularity conditions this gives that $\frac{k}{2}$-dimensional Hausdorff measure w.r.t. Kullback-Leibler distance is proportional to Jeffreys' prior (cf. Theorem 3.3).

2 Hausdorff measures with respect to a generalized distance

In this section we formulate some basic properties of the Hausdorff measure on a set X w.r.t. maps $F : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and $f : X \times X \longrightarrow \overline{\mathbb{R}}_+$. Most of the proofs in this section (cf. [3, Section 2.1]) can be given analogous to those of the corresponding propositions of the Hausdorff measure w.r.t. a metric (cf. [12, chap. I,II]).
Let X be a set, $F : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ an increasing map and $f : X \times X \longrightarrow \overline{\mathbb{R}}_+$ a map. Then $diam_f(A) := \sup\{f(x,y) \mid x, y \in A\}$ for sets $\emptyset \neq A \subset X$ and $diam_f(\emptyset) := 0$ defines the diameter of A w.r.t. f.
For $\delta > 0$ one can define outer measures on X by

$$\mathcal{H}_\delta^{F,\,f}(A) := \begin{cases} \inf\{\sum_{i=1}^\infty F(diam_f(A_i)) \mid A \subset \cup_{i=1}^\infty A_i,\ diam_f(A_i) \leq \delta,\ A_i \subset X\}, \text{ if such a} \\ \text{covering } (A_i)_{i \in \mathbb{N}} \text{ exists} \\ \infty, \text{ else} \end{cases}$$

and

$$\mathcal{H}^{F,\,f}(A) := \sup_{\delta > 0} \mathcal{H}_\delta^{F,\,f}(A)$$

(cf. [12, Theorem 15]).

Definition 2.1 *The outer measure $\mathcal{H}^{F,\,f}$ on X is called the Hausdorff measure on X w.r.t. (F, f). If $F(t) := t^\alpha$ for an $\alpha \geq 0$ we call $\mathcal{H}^{\alpha,\,f} := \mathcal{H}^{F,\,f}$ the α-dimensional Hausdorff measure on X w.r.t. f.*

If $\inf\{f(x,y) \mid x \in A,\ y \in E\} > 0$, the sets $A, E \subset X$ are called f-separated. In the same way as [12, Theorem 16] the following lemma can be proved.

Lemma 2.2 *Let X be a set, $f : X \times X \longrightarrow \overline{I\!R}_+$ a map with the property $f(x,y) = 0$ iff $x = y$. Then $\mathcal{H}^{F,\,f}(A \cup E) = \mathcal{H}^{F,\,f}(A) + \mathcal{H}^{F,\,f}(E)$ for all f-separated sets $A, E \subset X$.*

If (X,d) is a metric space and d-separated sets are also f-separated, then $\mathcal{H}^{F,\,f}$ is a metric measure in the sense of [12, Definition 14] by lemma 2.2 and therefore by [12, Theorem 19] the Borel sets are $\mathcal{H}^{F,\,f}$-measurable. So we get the following corollary.

Corollary 2.3 *Let (X,d) be a metric space and $f : X \times X \longrightarrow \overline{I\!R}_+$ a map with the properties*

(i) $f(x,y) = 0$ iff $x = y$

(ii) d-separated sets are f-separated.

Then the Borel sets in (X,d) are $\mathcal{H}^{F,\,f}$-measurable.

Remark 2.4 For strictly increasing continuous F with $F(0) = 0$ the Hausdorff measure w.r.t. (F,f) coincides with the 1-dimensional Hausdorff measure w.r.t. $F \circ f$. In particular the α-dimensional Hausdorff measure w.r.t. f coincides with the 1-dimensional Hausdorff measure w.r.t. f^α.

An important property of the Hausdorff measure is its behaviour under Hölder-continuous mappings. The following two lemmata can be proved analogously to the corresponding classical results (cf. [4, Prop. 2.2]).

Lemma 2.5 *Let X_1, X_2 be sets and let $f_i : X_i \times X_i \longrightarrow \overline{I\!R}_+$, $i = 1,2$ and $\Phi : X_1 \longrightarrow X_2$ be mappings. Assume that there are constants $c \geq 0, \gamma \geq 0$ such that $f_2(\Phi(x), \Phi(y)) \leq c f_1(x,y)^\gamma$ for all $x, y \in X_1$. Then $\mathcal{H}^{\alpha,\,f_2}(\Phi(X_1)) \leq c^\alpha \mathcal{H}^{\alpha\gamma,\,f_1}(X_1)$.*
If Φ is isometric, that is $f_2(\Phi(x), \Phi(y)) = f_1(x,y)$ for all $x, y \in X_1$, then $\mathcal{H}^{\alpha,\,f_2}(\Phi(X_1)) = \mathcal{H}^{\alpha,\,f_1}(X_1)$.

Lemma 2.6 *Let X_1, X_2 be sets and let $f_i : X_i \times X_i \longrightarrow \overline{I\!R}_+$, $i = 1,2$ and $\Phi : X_1 \longrightarrow X_2$ be mappings. For each $x \in X_1$ we assume that there is a $\delta_x > 0$ and a $L_x > 0$ such that $f_2(\Phi(x'), \Phi(x'')) \leq L_x f_1(x', x'')$ for all $x', x'' \in \{y \in X_1 \mid f_1(x,y) \leq \delta_x\}$. Then for $A \subset X_1$ one has $\mathcal{H}^{\alpha,\,f_2}(\Phi(A)) = 0$ if $\mathcal{H}^{\alpha,\,f_1}(A) = 0$.*

If X is a vector space (over the field $I\!R$) and $\|\cdot\|$ is a seminorm on X we denote by $B_{\|\cdot\|}$ the corresponding unit ball and by $\mathcal{H}^{k,\,\|\cdot\|}$ we denote the k-dimensional Hausdorff measure w.r.t. the pseudometric induced by the seminorm $\|\cdot\|$. If $\|\cdot\|$ is a norm on $I\!R^k$ (in this case we denote the norm by $|\cdot|$) the outer measure $\mathcal{H}^{k,\,|\cdot|}$ induces a translation invariant measure on the Borel sets, positive on each nonempty open set and finite on compact sets. By uniqueness of Haar measure on $I\!R^k$ and Borel regularity of both outer measures one concludes that $\mathcal{H}^{k,\,|\cdot|}$ is proportional to Lebesgue measure λ^k. The factor is given by $\frac{2^k}{\lambda^k(B_{|\cdot|})} = \mathcal{H}^{k,\,|\cdot|}(W)$ (cf. [8, Lemma 6], [3, Korollar 2.2.3]) where $W := \{x \in I\!R^k \mid 0 \leq x_i \leq 1 \text{ for all } 1 \leq i \leq k\}$ denotes the unit cube in $I\!R^k$. By $|\cdot|_2$ we denote Euclidean norm in Euclidean space $I\!R^k$.

Similar to the arguments in the first part of the proof of [8, Theorem 7] we get an area-formula for the Hausdorff measure $\mathcal{H}^{k,\,f}$ on Borel sets $A \subset I\!R^k$.

Proposition 2.7 *Let $A \subset I\!R^k$ be a Borel set and let $f : A \times A \longrightarrow \overline{R}_+$ be a map such that all Borel sets in A are $\mathcal{H}^{k,\,f}$-measurable. Further we assume for each $x \in A$ that there is a norm $|\cdot|_x$ on $I\!R^k$ such that for each $\varepsilon > 0$ there is a $\delta = \delta(x,\varepsilon) > 0$ with*

$$(1-\varepsilon)\left|x' - x''\right|_x \leq f(x', x'') \leq (1+\varepsilon)\left|x' - x''\right|_x$$

for all $x', x'' \in B(x,\delta) \cap A = \{y \in A \mid |x-y|_2 \leq \delta\}$.
Then

$$\mathcal{H}^{k,\,f}(E) = \int_E \frac{2^k}{\lambda^k(B_{|\cdot|_x})}\,\lambda^k(dx)$$

for all Borel sets $E \subset A$.

Proof: Let $\varepsilon > 0$ and $E \subset A$ be a Borel set. By Vitali's covering theorem we get a sequence $(x_i)_{i \in \mathbb{N}}$ in E and $0 < \delta_i \leq \delta(x_i, \varepsilon)$, $\quad i \in \mathbb{N}$, such that the balls $B(x_i, \delta_i)$ are disjoint and $\lambda^k\left(E \setminus \bigcup_{i=1}^{\infty} B(x_i, \delta_i)\right) = 0$. Moreover we have

$$\frac{1-\varepsilon}{1+\varepsilon}\left|x' - x''\right|_x \leq \left|x' - x''\right|_{x_i} \leq \frac{1+\varepsilon}{1-\varepsilon}\left|x' - x''\right|_x$$

for all $x \in E \cap B(x_i, \delta_i)$, $\quad x', x'' \in E \cap B(x_i, \delta_i) \cap B(x, \delta(x, \varepsilon))$. Thus

$$\begin{aligned}\mathcal{H}^{k,\,|\cdot|_{x_i}}(W)\,\lambda^k\left(E \cap B(x_i,\delta_i) \cap B(x,\delta(x,\varepsilon))\right) &= \mathcal{H}^{k,\,|\cdot|_{x_i}}\left(E \cap B(x_i,\delta_i) \cap B(x,\delta(x,\varepsilon))\right)\\ &\leq \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^k \mathcal{H}^{k,\,|\cdot|_x}(W)\,\lambda^k\left(E \cap B(x_i,\delta_i) \cap B(x,\delta(x,\varepsilon))\right)\end{aligned}$$

and

$$\begin{aligned}&\mathcal{H}^{k,\,|\cdot|_{x_i}}(W)\,\lambda^k\left(E \cap B(x_i,\delta_i) \cap B(x,\delta(x,\varepsilon))\right)\\ &\geq \left(\frac{1-\varepsilon}{1+\varepsilon}\right)^k \mathcal{H}^{k,\,|\cdot|_x}(W)\,\lambda^k\left(E \cap B(x_i,\delta_i) \cap B(x,\delta(x,\varepsilon))\right)\end{aligned}$$

for all $x \in E \cap B(x_i, \delta_i)$. Therefore we get

$$\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^k \mathcal{H}^{k,\,|\cdot|_x}(W) \leq \mathcal{H}^{k,\,|\cdot|_{x_i}}(W) \leq \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^k \mathcal{H}^{k,\,|\cdot|_x}(W)$$

for λ^k-almost all $x \in E \cap B(x_i, \delta_i)$. So using lemmata 2.2 up to 2.6 we have

$$\begin{aligned}\mathcal{H}^{k,\,f}(E) &= \mathcal{H}^{k,\,f}\left(\bigcup_{i=1}^{\infty}(E \cap B(x_i,\delta_i))\right)\\ &= \sum_{i=1}^{\infty}\mathcal{H}^{k,\,f}\left(E \cap B(x_i,\delta_i)\right)\\ &\leq (1+\varepsilon)^k \sum_{i=1}^{\infty}\mathcal{H}^{k,\,|\cdot|_{x_i}}(W)\,\lambda^k\left(E \cap B(x_i,\delta_i)\right)\\ &\leq (1+\varepsilon)^k \sum_{i=1}^{\infty}\int_{E \cap B(x_i,\delta_i)}\mathcal{H}^{k,\,|\cdot|_{x_i}}(W)\;\lambda^k(dx)\\ &\leq \frac{(1+\varepsilon)^{2k}}{(1-\varepsilon)^k}\sum_{i=1}^{\infty}\int_{E \cap B(x_i,\delta_i)}\mathcal{H}^{k,\,|\cdot|_x}(W)\;\lambda^k(dx)\\ &= \frac{(1+\varepsilon)^{2k}}{(1-\varepsilon)^k}\int_E \mathcal{H}^{k,\,|\cdot|_x}(W)\;\lambda^k(dx)\end{aligned}$$

and analogously

$$\frac{(1-\varepsilon)^{2k}}{(1+\varepsilon)^k}\int_E \mathcal{H}^{k,\,|\cdot|_x}(W)\;\lambda^k(dx) \leq \mathcal{H}^{k,\,f}(E)\,.$$

With ε declining to 0 this proves the lemma. $\circ$

3 The area formula for Hausdorff measures with respect to Hellinger and Kullback-Leibler distances

For this section let $\mathcal{F} := \left\{ P_\vartheta \mid \vartheta \in \Theta \subset I\!R^k \right\}$ be a family of probability measures on a measurable space $(\Omega, \mathcal{A})$ such that $P_\vartheta \neq P_{\vartheta'}$ whenever $\vartheta \neq \vartheta'$. We assume the family $\mathcal{F}$ to be dominated by a σ-finite measure μ and denote by $\{f_\vartheta \mid \vartheta \in \Theta\} \subset L^1(\mu)$ the corresponding density functions. If the map $\Phi : \Theta \longrightarrow L^2(\mu),\ \vartheta \mapsto \sqrt{f_\vartheta}$ is differentiable at $\vartheta_0 \in \Theta$ we call the matrix (w.r.t. the canonical algebraic basis) corresponding to the bilinear form $I(\vartheta_0) := 4D\Phi(\vartheta_0)^* D\Phi(\vartheta_0)$, the Fisher Information matrix of $\mathcal{F}$ at $\vartheta_0 \in \Theta$ and in this case $|\cdot|_\vartheta$ denotes the seminorm on $I\!R^k$ induced by the Fisher Information matrix $I(\vartheta)$, i.e. $|y|_\vartheta := \sqrt{\langle I(\vartheta)y, y\rangle},\ y \in I\!R^k$.
On the set $\Theta \times \Theta$ we consider the Hellinger distance $d(\vartheta, \vartheta') := \|\Phi(\vartheta) - \Phi(\vartheta')\|_2$ (cf. [5, I.7.6]) and the Kullback-Leibler distance $\mathcal{K}(\vartheta, \vartheta') := \int_\Omega f_\vartheta \ln \frac{f_\vartheta}{f_{\vartheta'}}\, d\mu$ (cf. [9]), respectively. In general for the Kullback-Leibler distance the triangle inequality failes so $\mathcal{K}$ is not a metric. One has $0 \leq \mathcal{K}(\vartheta, \vartheta') \leq \infty,\ \ \mathcal{K}(\vartheta, \vartheta') = 0$ iff $\vartheta = \vartheta'$ and $\mathcal{K}(\vartheta, \vartheta') < \infty$ implies that P_ϑ is absolutely continuous to $P_{\vartheta'}$.
In the sense of section 1 one can consider the Hausdorff measure on Θ w.r.t. Hellinger distance and Kullback-Leibler distance, respectively.
Let us consider the following regularity conditions

C1: The parameter set $\Theta \subset I\!R^k$ is open and $\Phi : \Theta \longrightarrow L^2(\mu),\ \vartheta \mapsto \sqrt{f_\vartheta}$ is an injective and locally Lipschitz-continuous function.

C2: The parameter set $\Theta \subset I\!R^k$ is open and Φ is injective and twice continuously differentiable. Further $f_\vartheta \ln \frac{f_{\vartheta'}}{f_\vartheta} \in L^1(\mu)$ for all $\vartheta, \vartheta' \in \Theta$ and for each $\vartheta \in \Theta$ the map $\Psi_\vartheta : \Theta \longrightarrow L^1(\mu),\ \vartheta' \mapsto f_\vartheta \ln \frac{f'_\vartheta}{f_\vartheta}$ is twice continuously differentiable in ϑ' and the second order derivative $D^2\Psi_\vartheta$ depends continuously on ϑ and ϑ'.

Remark: A real Banach space E has the Radon-Nikdym Property (RNP) if every function $f : [0,1] \longrightarrow E$ of bounded variation is differentiable λ^1-a.e. Every reflexive Banach space has the (RNP). By Aronszajn's theorem (cf. [1, Lemma2, chapter II] and [11, Theorem 6]) every locally Lipschitz-continuous function $f : I\!R^k \supset A \longrightarrow E,\ A$ open, is differentiable λ^k-a.e. Hence condition C1 implies Φ to be differentiable λ^k-a.e.

In the following we will show that under the regularity conditions C1 respectively C2 the k-dimensional Hausdorff measure w.r.t. Hellinger distance d and the $\frac{k}{2}$-dimensional Hausdorff meausre w.r.t. Kullback-Leibler distance $\mathcal{K}$ are proportional to Jeffreys' prior on the Borel σ-algebra of Θ, more precisely we have

$$\begin{aligned} \mathcal{H}^{k,\,d}(E \cap \Theta) &= \frac{1}{\omega_k} \int_{E\cap\Theta} \sqrt{\det I(\vartheta)}\, d\mu \qquad \text{and} \\ \mathcal{H}^{\frac{k}{2},\,\mathcal{K}}(E \cap \Theta) &= \frac{2^{\frac{k}{2}}}{\omega_k} \int_{E\cap\Theta} \sqrt{\det I(\vartheta)}\, d\mu \end{aligned}$$

for all Borel sets $E \subset I\!R^k$, where ω_k denotes the Lebesgue measure of the k-dimensional Euclidean unit ball and $I(\vartheta)$ denotes the Fisher Information matrix of the family $\{P_\vartheta \mid \vartheta \in \Theta\}$ at ϑ.
The following proposition shows that under the regularity condition C2 the squared Kullback-Leibler distance locally behaves as the seminorm $|\cdot|_\vartheta$ induced by the Fisher Information matrix.

Proposition 3.1 *Let the regularity conditions* C2 *be fulfilled. Then for each* $\vartheta_0 \in \Theta$ *and each* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that*

$$\left|\vartheta - \vartheta'\right|^2_{\vartheta_0} - \varepsilon \left|\vartheta - \vartheta'\right|^2 \leq 2\mathcal{K}(\vartheta, \vartheta') \leq \left|\vartheta - \vartheta'\right|^2_{\vartheta_0} + \varepsilon \left|\vartheta - \vartheta'\right|^2$$

for all $\vartheta, \vartheta' \in B(\vartheta_0, \delta)$.

Proof: Let $\vartheta, \vartheta' \in \Theta$ and $x, y \in I\!R^k$. Let $\varphi : \Theta \longrightarrow L^1(\mu)$, $\vartheta \mapsto f_\vartheta$. Then φ is differentiable with derivative $D\varphi(\vartheta) = 2\sqrt{f_\vartheta} D\Phi(\vartheta)$ at $\vartheta \in \Theta$ (cf. [3, Satz 3.1.5], [13, Lemma 5.3]) and $\int D\varphi(\vartheta)\, \lambda^k(d\vartheta) = 0 = \int x^t D^2\varphi(\vartheta) y\, \lambda^k(d\vartheta)$. Using standard subsequence arguments (cf. [2, Theorem 20.8]) the differentiability assumptions yield that one can get the derivatives of Ψ_ϑ by formal differentation, i. e. $D\Psi_\vartheta(\vartheta')y = f_\vartheta \frac{D\varphi(\vartheta')y}{f_{\vartheta'}}$ and $x^t D^2 \Psi_\vartheta(\vartheta') y = f_\vartheta \frac{x^t D^2 \varphi(\vartheta')y}{f_{\vartheta'}} - f_\vartheta \frac{x^t D\varphi(\vartheta')^* D\varphi(\vartheta')y}{f^2_{\vartheta'}}$ μ-almost everywhere and Taylor expansion of Ψ_ϑ at the point ϑ yields $-\mathcal{K}(\vartheta, \vartheta + \Delta) = -\frac{1}{2}\langle \Delta, I(\vartheta)\Delta\rangle + \int r(\vartheta, \Delta)\, d\mu$, where $r(\vartheta, \Delta)$ denotes the error of the second order Taylor expansion of Ψ_ϑ at ϑ.
Now let $\vartheta_0 \in \Theta$ and $\varepsilon > 0$. The Fisher Information matrix $I(\vartheta)$ depends continuously on ϑ, because Φ is continuously differentiable. Hence there is a $\delta_1 > 0$ such that the closed ball $B(\vartheta_0, \delta_1)$ is contained in Θ and $\sup_{\vartheta \in B(\vartheta_0, \delta_1)} \|I(\vartheta_0) - I(\vartheta)\| \le \frac{\varepsilon}{2}$. By the continuity of $(\vartheta, \vartheta') \mapsto D^2\Psi_\vartheta(\vartheta')$ there are $\delta_2 > 0, \delta_3 > 0$ such that $\sup_{\vartheta \in B(\vartheta_0, \delta_3)} |\int r(\vartheta, \Delta)\, d\mu| \le \frac{\varepsilon}{4} |\Delta|_2^2$ for all $|\Delta|_2 \le \delta_2$. Take $\delta := \frac{1}{2}\min\{\delta_1, \delta_2, \delta_3\}$, then we have $|\vartheta - \vartheta'|_2 \le 2\delta = \min\{\delta_1, \delta_2, \delta_3\}$ for all $\vartheta, \vartheta' \in B(\vartheta_0, \delta)$. Therefore

$$\begin{aligned} \left|2\mathcal{K}(\vartheta, \vartheta') - \langle \vartheta - \vartheta', I(\vartheta_0)(\vartheta - \vartheta')\rangle\right| &\le \|I(\vartheta) - I(\vartheta_0\| \left|\vartheta - \vartheta'\right|_2^2 + 2\left|\int r(\vartheta, \vartheta - \vartheta')\, d\mu\right| \\ &\le \varepsilon \left|\vartheta - \vartheta'\right|^2 . \end{aligned}$$

This proves the proposition. ∘

From the proof of theorem 3.3 below we separate the following lemma.

Lemma 3.2 *Let $\|\cdot\|$ be a seminorm on $I\!R^k$ which is not a norm and let $M > 0$ such that $\|x\| \le M|x|_2$ for all $x \in I\!R^k$. For $\tau > 0$ define $|x|_\tau := \|x\| + \tau|x|_2$, $x \in I\!R^k$. Then we have $\frac{1}{\lambda^k(B_{|\cdot|_\tau})} \le 2^{2k}(k-1)^{k-1}(\tau + M)^{k-1}\tau$ and therefore*

$$\mathcal{H}^{k,\, |\cdot|_\tau}(A) \le \tau\, 2^{2k}(k-1)^{k-1}(\tau + M)^{k-1}\lambda^k(A)$$

for all $A \subset I\!R^k$.

Proof: We can assume $k > 1$ because in the case $k = 1$ every seminorm which is not a norm is identically zero.
Take $0 \ne z \in I\!R^k$ such that $\|z\| = 0$ and $|z|_2 = 1$. Further take $\{z, e_2, \cdots, e_k\}$ an ONB of Euclidean space $I\!R^k$ and $x := \lambda_1 z + \sum_{j=2}^k \lambda_j e_j$ with $|\lambda_1| \le \frac{1}{2\tau}$ and $|\lambda_j| \le \frac{1}{2(k-1)(\tau+M)}$ for all $2 \le j \le k$. Then we have

$$\begin{aligned} |x|_\tau &= \|x\| + \tau|x|_2 \le \sum_{j=2}^k |\lambda_j|\, \|e_j\| + \tau\sum_{j=1}^k |\lambda_j| \le \tau|\lambda_1| + (\tau + M)\sum_{j=2}^k |\lambda_j| \\ &\le \frac{1}{2} + \frac{1}{2} = 1. \end{aligned}$$

Therefore we have $\frac{1}{\lambda^k(B_{|\cdot|_\tau})} \le 2^k\tau(k-1)^{k-1}(\tau + M)^{k-1}$. Because $\mathcal{H}^{k,\, |\cdot|_\tau} = \frac{2^k}{\lambda^k(B_{|\cdot|_\tau})}\lambda^k$ this proves the lemma. ∘

Now we show that the $\frac{k}{2}$-dimensional Hausdorff measure w.r.t. Kullback-Leibler distance and the k-dimensional Hausdorff measure w.r.t. Hellinger distance are proportional to Jeffreys' prior.

Theorem 3.3

(a) *If the regularity conditions* C1 *are satisfied we have $\mathcal{H}^{k,\, d}(\Theta') = \frac{1}{\omega_k}\int_{\Theta'} \sqrt{\det I(\vartheta)}\, d\mu$ for all Borel sets $\Theta' \subset \Theta \subset I\!R^k$.*

(b) *If the regularity conditions* C2 *are satisfied we have $\mathcal{H}^{\frac{k}{2},\, \mathcal{K}}(\Theta') = \frac{2^{\frac{k}{2}}}{\omega_k}\int_{\Theta'} \sqrt{\det I(\vartheta)}\, d\mu$ for all Borel sets $\Theta' \subset \Theta \subset I\!R^k$.*

Proof: The proof of part (a) is an immediate consequence of [8, Theorem 7] because in this case it is $\frac{2^k}{\lambda^k(B_{|\cdot|_\vartheta})} = \frac{\sqrt{\det I(\vartheta)}}{\omega_k}$.

Now let us prove (b). By assumption the map $\varphi : \Theta \longrightarrow L^1(\mu)$, $\vartheta \mapsto f_\vartheta$ is injective and continuous, hence by Kuratowski's theorem (cf. [10, Cor. 3.3]) Borel sets w.r.t. Euclidean metric are also Borel sets w.r.t. total variation metric. Further, by [7, Theorem 6.11] we have $\|f_\vartheta - f_{\vartheta'}\|_1 \leq \sqrt{2\mathcal{K}(\vartheta,\vartheta')}$ for all $\vartheta, \vartheta' \in \Theta$. Therefore corollary 2.3 implies that Borel subsets of Θ are measurable w.r.t. the Hausdorff measure $\mathcal{H}^{\frac{k}{2},\,\mathcal{K}}$.

First we consider the parameter set $\Theta_1 := \{\vartheta \in \Theta \mid \det I(\vartheta) > 0\}$. This is an open subset of $\mathbb{R}^k$ and for each $\vartheta \in \Theta$ $|y|_\vartheta := \sqrt{\langle y, I(\vartheta)y\rangle}$ defines a norm on $\mathbb{R}^k$. Hence by proposition 3.1 for each $\vartheta \in \Theta_1$ and $\varepsilon > 0$ there is a $\delta > 0$ such that

$$(1-\varepsilon)\left|\vartheta' - \vartheta''\right|_\vartheta \leq \sqrt{2}\sqrt{K(\vartheta',\vartheta'')} \leq (1+\varepsilon)\left|\vartheta' - \vartheta''\right|_\vartheta$$

for all $\vartheta', \vartheta'' \in B(\vartheta, \delta)$. Therefore we get by the area-formula 2.7

$$\mathcal{H}^{\frac{k}{2},\,\mathcal{K}}(E \cap \Theta_1) = \mathcal{H}^{k,\,\sqrt{\mathcal{K}}}(E \cap \Theta_1) = 2^{-\frac{k}{2}} \int \frac{2^k}{\lambda^k(B_{|\cdot|_\vartheta})}\, d\mu = \frac{2^{\frac{k}{2}}}{\omega_k} \int \sqrt{\det I(\vartheta)}\, d\mu.$$

Now we consider the parameter set $\Theta_2 := \{\vartheta \in \Theta \mid \det I(\vartheta) = 0\}$. As the complement of Θ_1 this is closed in Θ and therefore σ-compact. We will show $\mathcal{H}^{\frac{k}{2},\,\mathcal{K}}(E \cap \Theta_2) = 0$ for every Borel set $E \subset \mathbb{R}^k$.

First let $M \subset \Theta_2$ be compact. Then by proposition 3.1 for each $\varepsilon > 0$ there is a $l \in \mathbb{N}$ and parameters $\vartheta_1, \cdots, \vartheta_l \in M$ and $\delta_1, \cdots, \delta_l > 0$, such that $M \subset \cup_{i=1}^l B(\vartheta_i, \delta_i)$ and $\sqrt{2\mathcal{K}(\vartheta',\vartheta'')} \leq |\vartheta' - \vartheta''|_{\vartheta_i} + \varepsilon|\vartheta' - \vartheta''|_2$ for all $\vartheta', \vartheta'' \in B(\vartheta_i, \delta_i)$. Let us define $A_1 := M \cap B(\vartheta_1, \delta_1)$ and $A_i := M \cap B(\vartheta_i, \delta_i) \setminus \bigcup_{j=1}^{i-1} B(\vartheta_j, \delta_j)$ for $2 \leq j \leq l$, then by lemma 3.2 we have

$$\begin{aligned}
\mathcal{H}^{\frac{k}{2},\,\mathcal{K}}(E \cap M) &= \mathcal{H}^{\frac{k}{2},\,\mathcal{K}}\left(\cup_{i=1}^l E \cap A_i\right) = \sum_{i=1}^{l} \mathcal{H}^{\frac{k}{2},\,\mathcal{K}}(E \cap A_i) \\
&\leq \sum_{i=1}^{l} 2^{2k}(k-1)^{k-1}(\varepsilon + \sup_{\vartheta \in M} \|I(\vartheta)\|)^{k-1} \varepsilon \lambda^k(E \cap A_i) \\
&= 2^{2k}(k-1)^{k-1}(\varepsilon + \sup_{\vartheta \in M} \|I(\vartheta)\|)^{k-1} \varepsilon \lambda^k(E \cap M) \\
&\to 0 \quad \text{as } \varepsilon \to 0.
\end{aligned}$$

Because Θ_2 is σ-compact this yields $\mathcal{H}^{\frac{k}{2},\,\mathcal{K}}(E \cap \Theta_2) = 0$ for all Borel sets $E \subset \mathbb{R}^k$ and the theorem is proved. ◦

References

[1] N. Aronszajn. Differentiability of Lipschitz functions in Banach spaces. *Studia Math.*, 57:147–160, 1976.

[2] H. Bauer. *Maß und Integrationstheorie.* de Gruyter, Berlin - New York, 1990.

[3] W. Doster. *Zur Berechnung des Hausdorffmaßes von Familien von Wahrscheinlichkeitsmaßen.* PhD thesis, Universität Kaiserslautern, 1995.

[4] K.J. Falconer. *Fractal Geometry.* John Wiley & Sons, West Sussex, 1990.

[5] I.A. Ibragimov and R.Z. Has'minskii. *Statistical Estimation - Asymptotic Theory.* Springer Verlag, New York Heidelberg Berlin, 1981.

[6] H. Jeffreys. *Theory of Probability.* Clarendon Pr., Oxford, 3rd edition, 1983.

[7] J.H.B. Kempermann. On the optimum rate of transmitting information. In *Probability and Information Theory*, pages 126 – 169. Lecture Notes in Mathematics, Springer Verlag, 1967.

[8] B. Kirchheim. Rectifiable metric spaces: Local structure and regularity of the Hausdorff measure. *Proc. Amer. Math. Soc.*, (121):113 – 123, 1994.

[9] S. Kullback. *Information Theory and Statistics.* Dover Publications, Dover, 1968. Dover Auflage.

[10] K.R. Parthasarathy. *Probability measures on metric spaces.* Academic Press, New York, 1967.

[11] R.R Phelps. Gaussian null sets and differentiability of Lipschitz maps on Banach spaces. *Pac. J. Math.*, 77:523–531, 1978.

[12] C.A. Rogers. *Hausdorff Measures.* Cambridge University Press, 1970.

[13] O.G. Smolyanov and H. v. Weizsäcker. Differential families of measures. *Journal of Functional Analysis*, 118:454–476, 1993.