Differentiability of Statistical Experiments

Helmut Strasser

Received: Revised version:

Abstract: Let $E = (P_\theta)$ be a dominated experiment with an open Euclidean parameter space and densities (f_θ). The experiment is said to be continuously L^2–differentiable if the family $(\sqrt{f_\theta})$ is continuously L^2–differentiable. The following assertions are proved:
(1) The experiment E is continuously L^2–differentiable iff the family (f_θ) is continuously L^1–differentiable and Fisher's information function is continuous.

(2) Suppose that the experiment E is continuously L^2–differentiable and the experiment F is less informative than E. Then F is continuously L^2–differentiable, too.

1 Introduction

This paper is about differentiability concepts for families of probability measures. It is true that this topic is not a very recent one and almost everything has already been said about it. Nevertheless, reading the recent textbook literature gives the impression that some remarks might still be of interest. Such a remark is the subject of this paper.

Suppose we are given a family (P_θ) of probability measures on a σ–finite measure space $(\Omega, \mathcal{A}, \mu)$. The parameter space Θ is an open subset of $\mathbb{R}^d$. The standard theory of asymptotic statistics is based on the condition that the family (P_θ) is regular. The mathematical definition of regularity has been fixed by many authors in similar but different ways. After the first working concepts by Wald, [12], and Cramer, [2], the most important achievements are due to LeCam, [6], and Hájek, [4]. The state of the art is presented by Bickel, Klaassen, Ritov and Wellner, [1].

Let (f_θ) be the μ–densities of the family (P_θ). Then it is common knowledge that the basic asymptotic expansion argument (local asymptotic normality) works if the mapping $\theta \mapsto \sqrt{f_\theta}$ is differentiable as a mapping from Θ to the Banach space $L^2(\mu)$. Since checking this condition for a particular family is sometimes not an easy matter many authors tried to give sufficient conditions which can be stated in terms of the densities themselves instead of their square roots. A necessary and sufficient condition has been obtained by Pfanzagl, [9], Lemma 1.2.17.

[1] *AMS 1991 subject classifications.* Primary: 62A20, 62B15; secondary: 62F05, 62F12

[2] *Key words and phrases.* Regularity, differentiability in quadratic mean, sufficiency, equivalence of experiments.

A major question is whether L^2–differentiability of $(\sqrt{f_\theta})$ can be equivalently stated by simple smoothness properties of the densities. Many authors have discussed this problem. Recent examples are Witting, [13], Satz 1.194, LeCam and Yang, [8], 6.2, and Bickel, Klaassen, Ritov and Wellner, [1], section 2.1, Proposition 1. Mathematically, the most sophisticated treatment of this problem is given by LeCam, [7], pp. 585, ff. As it is shown by LeCam, mere L^2–differentiability of $(\sqrt{f_\theta})$ does not imply continuity or smoothness properties of the densities (cf. LeCam, [7], p. 589, f.).

In the present paper we would like to show that things are far easier if we consider continuous L^2–differentiability instead of mere L^2–differentiability of $(\sqrt{f_\theta})$. The first main assertion we will prove is as follows (Theorem (2.8)):

Suppose that that Fisher's information function is finite and continuous. Then continuous L^1–differentiability of (f_θ) is equivalent with continuous L^2–differentiability of $(\sqrt{f_\theta})$.

Of course, this assertion needs some comments.

It is well–known that L^2–differentiability of $(\sqrt{f_\theta})$ implies L^1–differentiability of (f_θ). If anything is interesting in our assertion then it is the converse part. Usually L^1–differentiability of densities is easy to check and thus, our result could lead to some simplification of proofs for particular models.

In the proof of the assertion we almost need not use any new ideas. The major parts of the proof could be given by quoting known arguments and putting them together in a slightly different way than usual. For the convenience of the reader, however, we will give a complete proof with extensive references to the origin of the arguments (section 4).

The outline of our argument is as follows. For the verification of continuous differentiability it is sufficient to verify existence and continuity of the partial derivatives. This latter problem is a one–dimensional problem and can be treated via absolute continuity. Absolute continuity has been applied by Hájek, [4], in order to prove L^2–differentiability of $(\sqrt{f_\theta})$. Our basic idea is simply to show that continuous L^1–differentiability of the densities implies absolute continuity of the densities along straight lines. Then Hájek's method gives the result.

Is it worth considering continuous differentiability ? This question can be discussed on different levels of sophistication.

Let us first consider a simple technical point of view. Certainly, there are situations which are interesting from the theoretical point of view where mere differentiability holds but continuity of the derivative is not fulfilled. Examples mentioned by LeCam, [7], p. 589, are of that type. However, the regular case, as it is defined e.g. by Bickel, Klaassen, Ritov and Wellner, [1], section 2.1, Definition 2, requires continuity of the derivative. As a matter of fact many authors suggest sufficient conditions for L^2–differentiability of $(\sqrt{f_\theta})$ which imply even continuous differentiability. This is true of Witting, [13], Satz, 1.194, Pfanzagl and Wefelmeyer, [10], Proposition 2.2.1, LeCam and Yang, [8], p. 102, Bickel, Klaassen, Ritov and Wellner, [1], section 2.1, Proposition 1. Each of these sets of sufficient conditions seems not to be necessary for continuous differentiability, or at least the necessity is not considered.

Thus, our result gives a further set of conditions for continuous L^2–differentiability of $(\sqrt{f_\theta})$ which has the advantage of being even necessary.

Now let us turn to a more philosphical point of view. A property stated for a statistical experiment is statistically meaningful only if it is shared by all equivalent experiments (in the sense of Blackwell and LeCam). Therefore LeCam, [7], chapter 17, section 3, suggests to define regularity properties in terms of the Hellinger distances only (condition (l)). Such properties are shared by equivalent experiments.

Applying our criterion we are able to prove that continuous L^2–differentiability is also shared by equivalent experiments. In fact, we are able to prove the hereditary property for continuous L^2–differentiability. This is our second main result (Theorem (3.3)):

Suppose that the experiment E is more informative than the experiment F. If E is continuously L^2–differentiable then F is continuously L^2–differentiable, too.

Thus, continuous differentiability is a smoothness property of experiment types and thus fulfils the requirements of a statistically meaningful concept.

Probably, our criterion will not be able to cover any interesting example which has not been treated previously by other methods. However, some results could be proved in a more direct manner if our criterion is applied. E.g., this can be seen from the proof of the hereditary property. Moreover, applying the main result of this paper we will easily obtain known results concerning families of induced measures and of mixtures.

2 The criterion

In this section we will state and explain a criterion for continuous L^2–differentiability. Let us begin with some preliminary remarks.

Let $\Theta \subseteq \mathbb{R}^d$ be an open set and let $(\Omega, \mathcal{A}, \mu)$ be a σ–finite measure space. For every $\theta \in \Theta$ let P_θ be a μ–continuous probability measure with density

$$f_\theta = f(.,\theta) = \frac{dP_\theta}{d\mu}$$

and denote the square root of the densities by $s_\theta := \sqrt{f_\theta}$. We consider the statistical experiment $E = (\Omega, \mathcal{A}, (P_\theta, \theta \in \Theta))$.

(2.1) Definition *The experiment E is continuously L^1–differentiable if the function $\theta \to f_\theta$ is continuously differentiable as a function from Θ to $L^1(\Omega, \mathcal{A}, \mu)$.*

(2.2) Definition *The experiment E is continuously L^2–differentiable if the function $\theta \to s_\theta$ is continuously differentiable as a function from Θ to $L^2(\Omega, \mathcal{A}, \mu)$.*

It is obvious that the notions of L^1– or L^2–differentiability do not depend on the choice of the dominating measure μ.

If E is L^1–differentiable then the family (f_θ) is differentiable in μ–measure, too, and the L^1–derivative and the μ–derivative coincide. If it exists we will denote the μ–derivative of (f_θ) at θ by $\dot{f}_\theta$. If $d > 1$ we consider $\dot{f}_\theta$ as a row vector of functions. Similarly, if it exists we will denote the L^2–derivative of (s_θ) at θ by $\dot{s}_\theta$. Again, if $d > 1$ we consider $\dot{s}_\theta$ as a row vector of functions. Both notions of derivatives depend on the dominating measure μ.

A preliminary and well–known result is that L^2–differentiability implies L^1–differentiability. For completeness this is proved in Section 4.

(2.3) Theorem *Let $\Theta \subseteq \mathbb{R}^d$ be open. Suppose that E is L^2–differentiable at θ with derivative $\dot{s}_\theta$. Then E is L^1–differentiable at θ with derivative $\dot{f}_\theta = 2s_\theta\dot{s}_\theta$.*

We are especially interested in continuous differentiablitiy.

(2.4) Corollary *Let $\Theta \subseteq \mathbb{R}^d$ be open. Suppose that E is continuously L^2–differentiable. Then E is continuously L^1–differentiable.*

Proof: Apply Theorem (2.3) and Lemma (5.6). □

If we only know that the family (f_θ) is differentiable in μ–measure then we have to start with the derivative $\dot{f}_\theta$. Let us define what we mean by square root and loglikelihood derivatives. Define

$$\dot{s}(\omega,\theta) := \begin{cases} \dfrac{\dot{f}(\omega,\theta)}{2s(\omega,\theta)} & \text{where } f(\omega,\theta) > 0, \\ 0 & \text{elsewhere.} \end{cases} \tag{2.5}$$

and

$$\dot{l}(\omega,\theta) := \begin{cases} \dfrac{\dot{f}(\omega,\theta)}{f(\omega,\theta)} & \text{where } f(\omega,\theta) > 0, \\ 0 & \text{elsewhere.} \end{cases} \tag{2.6}$$

The functions $\dot{s}_\theta$ are the square root derivatives and the functions $\dot{l}_\theta$ are the loglikelihood derivatives. In cases where the densities are pointwise smooth and positive our definitions coincide with the usual notions. It should be noted that the loglikelihood derivatives $\dot{l}_\theta$ do not depend on the dominating measure μ. Thus, the loglikelihood derivatives are natural candidates for the notion of a derivative of the experiment $E = (P_\theta)$. Sometimes they are called tangent vectors of the experiment.

If the family of densities (f_θ) is differentiable in μ–measure then Fisher's information function can be defined.

(2.7) Definition *Suppose that (f_θ) is differentiable in μ–measure. Then*

$$I(\theta) := \int \dot{l}_\theta' \cdot \dot{l}_\theta \, dP_\theta = 4\int \dot{s}_\theta' \cdot \dot{s}_\theta \, d\mu$$

is called Fisher's information at θ.

As matter of fact Fisher's information is statistically meaningful only if the experiment E is continuously L^2–differentiable. If E is continuously L^2–differentiable then Fisher's information is a continuous function and satisfies

$$h'I(\theta)h = 4 \lim_{t\to 0} \frac{1}{t^2} \int (s_{\theta+th} - s_\theta)^2 \, d\mu.$$

Hence, Fisher's information function determines the local structure of the Hellinger metric of the experiment E.

Now we are in a position to state the converse of Corollary (2.4). The result is as follows.

(2.8) THEOREM *Suppose that E is continuously L^1–differentiable with derivative $(\dot{f}_\theta)$. If Fisher's information function is finite and continuous then E is continuously L^2–differentiable with derivative $\dot{s}_\theta$ defined by (2.5).*

This theorem is proved in section 4.

It is sometimes inconvenient to check L^1–differentiability. If Fisher's information is continuous then it is possible to state equivalent conditions which can be checked pointwise.

(2.9) THEOREM *Suppose that the family of densities (f_θ) is differentiable in μ–measure and that the derivatives $\dot{f}_\theta$ are μ–continuous. If Fisher's information is finite and continuous then E is continuously L^1– (and hence also L^2–) differentiable on Θ iff*

$$f_{\theta+h} - f_\theta = \int_0^1 \dot{f}_{\theta+th} \cdot h \, dt \quad \mu\text{–a.e.} \tag{2.10}$$

whenever the interval between θ and $\theta + h$ is contained in Θ.

This is also proved in Section 4. It should be noted that in (2.10) the exceptional set may depend on θ and on h.

3 Applications

In this section we will show that our criterion for L^2–differentiability leads to very simple proofs of basic facts.

A first application is devoted to families of induced probability measures. The result is well–known. Recent references are Witting, [13], Satz 1.193, Bickel, Klaassen, Ritov, Wellner, [1], Appendix A.5, Proposition 5. We provide a short proof in order to illustrate the application of our criterion.

(3.1) THEOREM *Suppose that the family (P_θ) is continuously L^2–differentiable. Let $\mathcal{C} \subseteq \mathcal{A}$ be a sub–σ–field. Then the family $(P_\theta|\mathcal{C})$ is continuously L^2–differentiable, too.*

Proof: Without loss of generality we assume for convenience that the dominating measure μ is a probability measure. Moreover, it is sufficient to consider the one–dimensional case. The family $(P_\theta|\mathcal{C})$ has the μ–densities $(E_\mu(f_\theta|\mathcal{C}))$. Let $\dot{f}_\theta$ be the L^1–derivative of (P_θ). Since (P_θ) is continuously L^1–differentiable it is obvious that $(P_\theta|\mathcal{C})$ is L^1–differentiable with derivative $E_\mu(\dot{f}_\theta|\mathcal{C})$ and that the derivatives are L^1–continuous.
Let $M_\theta = \{E_\mu(f_\theta|\mathcal{C}) > 0\}$. Since

$$\int_{M'_\theta} f_\theta \, d\mu = \int_{M'_\theta} E_\mu(f_\theta|\mathcal{C}) = 0$$

we have $M'_\theta \subseteq \{f_\theta = 0\}$ μ–a.e.
It remains to be checked that for the experiment $(P_\theta|\mathcal{C})$ Fisher's information function is continuous. Let $\dot{s}_\theta$ be the L^2–derivative of (P_θ) and define by

$$\dot{t}_\theta := \begin{cases} \dfrac{E_\mu(\dot{f}_\theta|\mathcal{C})^2}{2\sqrt{E_\mu(f_\theta|\mathcal{C})}} & \text{on } M_\theta \\ 0 & \text{elsewhere} \end{cases}$$

the analogon of (2.5) for the experiment $(P_\theta|\mathcal{C})$. For this experiment Fisher's information is

$$I_{\mathcal{C}}(\theta) = 4 \int \dot{t}_\theta^2 \, d\mu.$$

We have

$$\dot{t}_\theta^2 = \Big(\frac{E_\mu(\dot{s}_\theta\sqrt{f_\theta}|\mathcal{C})}{\sqrt{E_\mu(f_\theta|\mathcal{C})}}\Big)^2 \leq E_\mu(\dot{s}_\theta^2|\mathcal{C}) \quad \text{on } M_\theta.$$

Hence $(\dot{t}_\theta)$ is uniformly L^2-integrable. It remains to be shown that $\theta \mapsto \dot{t}_\theta$ is μ–continuous. Let $\theta_n \to \theta$. Then $\dot{t}_{\theta_n} \xrightarrow{\mu} \dot{t}_\theta$ is clear on M_θ since the conditional expectation preserves the continuity properties of the experiment $(P_\theta|\mathcal{C})$. But it is also true on M'_θ since

$$\int_{M'_\theta} E_\mu(\dot{s}_{\theta_n}^2|\mathcal{C}) \, d\mu \leq \int_{M'_\theta} \dot{s}_{\theta_n}^2 \, d\mu \leq \int_{f_\theta=0} \dot{s}_{\theta_n}^2 \, d\mu \;\to\; 0.$$

□

The preceding assertion has an interesting decision theoretic consequence. A major drawback of the usual Cramèr–Wald regularity conditions is their dependence on the particular representation of the experiment. We will show that continuous L^2–differentiablity is a regularity condition which is shared by all experiments of an equivalence class. We will even prove the stronger assertion that continuous L^2–differentiability is an hereditary property. This means that continuous L^2–differentiability is shared by all less informative experiments.

We need a preliminary decision theoretic lemma.

(3.2) LEMMA *Let E and F be dominated experiments for a parameter space Θ. Then E is more informative than F ($E \supseteq F$) iff there exists an experiment $G = (\Omega, \mathcal{A}, (R_\theta))$ and sub–σ–fields $\mathcal{C}_1 \subseteq \mathcal{A}$ and $\mathcal{C}_2 \subseteq \mathcal{A}$ satisfying the following conditions:*

(1) $E = G\,|\,\mathcal{C}_1$ and $\mathcal{C}_1$ is G–sufficient.

(2) $F = G\,|\,\mathcal{C}_2$.

Proof: Assume that conditions (1) and (2) are satisfied. Then condition (1) implies that E is equivalent to G. Condition (2) implies that G is more informative than F. Hence, E is more informative than F.

Now, assume conversely that E is more informative than F. Let $E = (\Omega_1, \mathcal{C}_1, (P_\theta))$ and $F = (\Omega_2, \mathcal{C}_2, (Q_\theta))$. Let $\mu \sim (P_\theta)$ and $\nu \sim (Q_\theta)$. Then, by the randomization criterion there is a stochastic operator $T : L^1(\mu) \to L^1(\nu)$ such that $TP_\theta = Q_\theta$ for all $\theta \in \Theta$. Define

$$R_\theta(C_1 \times C_2) := \int_{C_2} T\Big(\frac{dP_\theta}{d\mu} 1_{C_1}\Big)\, d\nu, \quad C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2.$$

This defines probability measures R_θ on $\mathcal{C}_1 \otimes \mathcal{C}_2$. Define the experiment $G := (\Omega_1 \times \Omega_2, \mathcal{C}_1 \otimes \mathcal{C}_2, (R_\theta))$. It is easy to see that $R_\theta|\mathcal{C}_1 = P_\theta$ and $R_\theta|\mathcal{C}_2 = Q_\theta$.

It remains to show that $\mathcal{C}_1$ is G–sufficient. Let T^* be the adjoint Markov operator of T. We will show that

$$E_{R_\theta}(f|\mathcal{C}_1) = T^*(f) \quad R_\theta\text{–a.e.}$$

for any bounded $\mathcal{C}_1 \otimes \mathcal{C}_2$–measurable function on $\Omega_1 \times \Omega_2$. For this we note that

$$\begin{aligned}\int_{C_1} T^*(f(\omega_1,.))\, R_\theta(d\omega_1, d\omega_2) &= \int_{C_1} T^*(f(\omega_1,.))\, P_\theta(d\omega_1) \\ &= \int_{C_1} f(\omega_1,.)\, R_\theta(d\omega_1, d\omega_2)\end{aligned}$$

by definition of R_θ. □

The preceding lemma leads to a simple proof of the hereditary property.

(3.3) THEOREM *Suppose that the experiment $E = (P_\theta)$ is more informative than the experiment $F = (Q_\theta)$. If E is continuously L^2–differentiable then the same is true of F.*

Proof: Let G be an experiment such that conditions (1) and (2) of Lemma (3.2) are satisfied. If we show that G is continuously L^2–differentiable then the assertion is proved by Theorem (3.1).

The experiment E is continuously L^2–differentiable and $E = G\,|\,\mathcal{C}_1$ for a G–sufficient σ–field $\mathcal{C}_1$. By the Halmos–Savage factorization theorem the densities of E and G are proportional up to a nonnegative measurable function which is independent of θ. Thus it is easy to see that continuous L^2–differentiability carries over from E to G. $\square$

Our second application deals with families of mixtures.

Let us fix the notation. Let $(\Lambda, \mathcal{C}, ?)$ be a probability space. For every $\eta \in \Lambda$ let $(P_{\theta,\eta})_{\theta\in\Theta}$ be a continuously L^2–differentiable family. The μ–densities of $P_{\theta,\eta}$ are denoted by $f_\theta(.,\eta)$. It is assumed that $(\omega,\eta) \mapsto f_\theta(\omega,\eta)$ is $\mathcal{A}\otimes\mathcal{C}$–measurable for every $\theta \in \Theta$.

For the following theorem we define

$$Q_\theta(A\times C) = \int_C P_{\theta,\eta}(A)\,?\,(d\eta) \quad \text{if } A \in \mathcal{A} \text{ and } C \in \mathcal{C}.$$

(3.4) THEOREM *For every $\eta \in \Lambda$ let $(P_{\theta,\eta})_{\theta\in\Theta}$ be a continuously L^2–differentiable family with Fisher's information $I(\theta,\eta)$. If the family of functions $(I(\theta,.))_{\theta\in\Theta}$ is uniformly ?–integrable then the family of mixtures (Q_θ) is continuously L^2–differentiable.*

Proof: In this case it is inconvenient to check L^1–differentiability. We will therefore apply Theorem (2.9).

For every $\eta \in \Lambda$ let $\dot{f}_\theta(.,\eta)$ be the L^1-derivative of $(P_{\theta,\eta})_{\theta\in\Theta}$. It is clear that the densities f_θ are differentiable in $\mu\otimes ?$–measure with $\mu\otimes ?$–continuous derivative $\dot{f}_\theta$. For every $\eta\in\Lambda$ let $\dot{s}_\theta(.,\eta)$ be the L^2-derivative of $(P_{\theta,\eta})_{\theta\in\Theta}$. Then the family (Q_θ) has Fisher's information $I(\theta) = \int ||\dot{s}_\theta(.,\eta)||_2^2\,?\,(d\eta)$. It is then clear from the assumptions that $\theta \mapsto I(\theta)$ is continuous.

It remains to check condition (2.10). For every $\eta\in\Lambda$ we have by Lemma (5.2)

$$f_{\theta+h}(.,\eta) - f_\theta(.,\eta) = \int_0^1 \dot{f}_{\theta+th}(.,\eta)\cdot h\,dt \quad \mu\text{–a.e.}$$

whenever the interval between θ and $\theta+h$ is contained in Θ. This implies

$$f_{\theta+h} - f_\theta = \int_0^1 \dot{f}_{\theta+th}\cdot h\,dt \quad \mu\otimes ?\text{–a.e.}$$

whenever the interval between θ and $\theta+h$ is contained in Θ. Hence the assertion. $\square$

Now we are ready to reproduce a known result which is proved in Bickel, Klaassen, Ritov, Wellner, [1], 4.5, Theorem 2. Let

$$Q'_\theta(A) = Q_\theta(A\times\Lambda), \quad A\in\mathcal{A}.$$

(3.5) COROLLARY *Suppose that the assumptions of Theorem (3.4) are satisfied. Then the family (Q'_θ) is continuously L^2–differentiable.*

Proof: Apply Theorems (3.4) and (3.1). $\square$

4 Proofs of the main results

As a first result we show that L^2–differentiability implies L^1–differentiability. This is well–known (cf. Bickel, Klaassen, Ritov and Wellner, [1], A.5, Prop. 3, or Witting, [13], Satz 1.190).

Proof: (of Theorem (2.3))
The derivative $\dot{s}_\theta$ is a row vector of functions in $L^2(\Omega, \mathcal{A}, \mu)$ such that

$$\lim_{|h|\to 0} \frac{1}{|h|^2} \int (\sqrt{f_{\theta+h}} - \sqrt{f_\theta} - \dot{s}_\theta \cdot h)^2 \, d\mu = 0.$$

We apply Lemma (5.1). It follows that

$$\frac{1}{|h|} \int |f_{\theta+h} - f_\theta - 2\sqrt{f_\theta}\, \dot{s}_\theta \cdot h| \, d\mu$$

$$\leq \frac{1}{|h|} \int (\sqrt{f_{\theta+h}} - \sqrt{f_\theta})^2 \, d\mu + \frac{2}{|h|} \int |\sqrt{f_{\theta+h}} - \sqrt{f_\theta} - \dot{s}_\theta \cdot h| \sqrt{f_\theta} \, d\mu$$

$$\leq \frac{1}{|h|} \int (\sqrt{f_{\theta+h}} - \sqrt{f_\theta})^2 \, d\mu + \frac{2}{|h|} \Big(\int (\sqrt{f_{\theta+h}} - \sqrt{f_\theta} - \dot{s}_\theta \cdot h)^2 \, d\mu \Big)^{1/2}.$$

This proves the assertion. □

It is our main goal to prove Theorem (2.8). The essential steps (4.1) to (4.5) are carried out for one–dimensional parameters.

The first assertion simply states that continuous differentiability implies absolute continuity. This is nothing else than the Newton–Leibniz fundamental theorem of calculus for Banach space–valued functions. It is also applied in a similar way by Bickel, Klaassen, Ritov and Wellner, [1], A.5, Prop. 3.

(4.1) Lemma *Let $\Theta = (\alpha, \beta)$ and suppose that E is continuously L^1–differentiable. Then*

$$\Big|\Big| f_\theta - f_{\theta_0} - \int_{\theta_0}^{\theta} \dot{f}_t \, dt \Big|\Big|_1 = 0 \quad \textit{for all } \theta_0, \theta \in \Theta..$$

Proof: Since $\theta \to \dot{f}_\theta$ is L^1–continuous it is clear that

$$\lim_{h\to 0} \frac{1}{h} \Big|\Big| \int_\theta^{\theta+h} \dot{f}_t \, dt - h\dot{f}_\theta \Big|\Big|_1 \leq \lim_{h\to 0} \frac{1}{h} \int_\theta^{\theta+h} ||\dot{f}_t - \dot{f}_\theta||_1 \, dt = 0.$$

Choose $\theta_0 \in \Theta$ and keep it fixed. Let

$$G(\theta) := f_\theta - f_{\theta_0} - \int_{\theta_0}^{\theta} \dot{f}_t \, dt.$$

Since E is L^1–differentiable and $\theta \to \dot{f}_\theta$ is L^1–continuous it follows that

$$\lim_{h\to 0} \frac{1}{h} \int \Big| G(\theta+h) - G(\theta) \Big| \, d\mu = 0.$$

Hence, G is differentiable on Θ and its derivative is identically zero. It follows that G is constant on Θ which is only possible if $||G(\theta)||_1 = 0$ on Θ (apply Dieudonné, [3], (8.6.1)). □

If E is L^1–differentiable then the family (P_θ) is continuous and the densities (f_θ) can be chosen as a separable family of functions (cf. Strasser, [11], Theorem 4.8). We are going to show that in this case separable densities are even pathwise absolutely continuous. This will close the gap between continuous L^1–differentiability and Hájek's set of conditions.

(4.2) LEMMA *Let $\Theta = (\alpha, \beta)$ and suppose that E is continuously L^1–differentiable. If the densities (f_θ) are a separable family then there is a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ and*

$$f(\omega, \theta) - f(\omega, \theta_0) = \int_{\theta_0}^{\theta} \dot{f}_t(\omega) \, dt \quad \text{for all } \theta_0, \theta \in \Theta \text{ and } \omega \notin N.$$

Proof: Let S be a countable separant of (f_θ) and $N \in \mathcal{A}$ the pertaining exceptional set with $\mu(N) = 0$. By Lemma (5.4) we may choose N even such that

$$\int_{\theta_0}^{\theta} |\dot{f}_t(\omega)| \, dt < \infty \quad \text{if } \theta_0, \theta \in (\alpha, \beta) \text{ and } \omega \notin N.$$

We apply Lemma (4.1) to obtain a set $N_1 \in \mathcal{A}$, $\mu(N_1) = 0$, such that

$$f(\omega, \theta) - f(\omega, \theta_0) = \int_{\theta_0}^{\theta} \dot{f}_t(\omega) \, dt \quad \text{for all } \theta_0, \theta \in S \text{ and } \omega \notin N_1.$$

If $\omega \notin N \cup N_1$ then separability implies

$$f(\omega, \theta) - f(\omega, \theta_0) = \int_{\theta_0}^{\theta} \dot{f}_t(\omega) \, dt \quad \text{for all } \theta_0, \theta \in \Theta.$$

□

Recall Definition (2.7). The following lemma is the turning point where we begin to consider differentiability of the square roots of the densities. The subsequent arguments are due to Hájek, [4]. They are repeated for the reader's convenience.

(4.3) LEMMA *Let $\Theta = (\alpha, \beta)$ and suppose that E is continuously L^1–differentiable. Assume further that $\theta \mapsto ||\dot{s}_\theta||_2$ is finite and continuous. If the densities (f_θ) are a separable family then there is a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ and*

$$\sqrt{f(\omega,\theta)} - \sqrt{f(\omega,\theta_0)} = \int_{\theta_0}^{\theta} \dot{s}_t(\omega)\,dt \quad \textit{for all } \theta_0, \theta \in \Theta \textit{ and } \omega \notin N.$$

Proof: The proof is based on Lemma (5.3). From Lemma (4.2) and Lemma (5.5) it follows that there is a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ and such that for every $\omega \notin N$ the function $\theta \mapsto f(\omega, \theta)$ satisfies the conditions of Lemma (5.3). □

(4.4) COROLLARY *Suppose that the assumptions of Lemma (4.3) are satisfied. Then*

$$\limsup_{h\to 0} \int \Big(\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h}\Big)^2 d\mu \le \int \dot{s}_\theta^2\,d\mu.$$

Proof: The following chain of expressions is a well-known argument which goes back to LeCam, [6]:

$$\begin{aligned}
\int \Big(\sqrt{f_{\theta+h}} - \sqrt{f_\theta}\Big)^2 d\mu &= \int \Big(\int_0^h \dot{s}_{\theta+t}\,dt\Big)^2 d\mu \\
&= \int \int_0^h \dot{s}_{\theta+t_1}\,dt_1 \int_0^h \dot{s}_{\theta+t_2}\,dt_2\,d\mu \\
&= \int_0^h \int_0^h \int \dot{s}_{\theta+t_1}\dot{s}_{\theta+t_2}\,d\mu\,dt_1\,dt_2 \\
&\le \int_0^h \int_0^h ||\dot{s}_{\theta+t_1}||_2\,||\dot{s}_{\theta+t_2}||_2\,dt_1\,dt_2 \\
&= \Big(\int_0^h ||\dot{s}_{\theta+t}||_2\,dt\Big)^2 \;\le\; |h| \int_0^h ||\dot{s}_{\theta+t}||_2^2\,dt
\end{aligned}$$

By $h \to 0$ we obtain the assertion. □

Now we are in the position to prove the main result.

(4.5) THEOREM *Let $\Theta = (\alpha, \beta)$ and suppose that E is continuously L^1–differentiable. Assume further that $\theta \mapsto I(\theta) = ||\dot{s}_\theta||_2^2$ is continuous. Then E is continuously L^2–differentiable.*

Proof: It is sufficient to prove the assertion for separable densities. Hence we may apply Lemma (4.3).

Let $M_\theta = \{f_\theta > 0\}$. On M_θ we have

$$\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h} = \frac{f_{\theta+h} - f_\theta}{h} \frac{1}{\sqrt{f_{\theta+h}} + \sqrt{f_\theta}}.$$

This implies that

$$\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h} \xrightarrow{\mu} \dot{s}_\theta \quad \text{on } M_\theta.$$

From Corollary (4.4) we obtain

$$\limsup_{h\to 0} \int_{M_\theta} \Big(\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h}\Big)^2 d\mu \le \int_{M_\theta} \dot{s}_\theta^2 \, d\mu = \int \dot{s}_\theta^2 \, d\mu. \tag{4.6}$$

The Lemma of Scheffé gives us

$$\lim_{h\to 0} \int_{M_\theta} \Big(\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h} - \dot{s}_\theta\Big)^2 d\mu = 0.$$

Now, it is clear that in (4.6) even equality holds which together with Corollary (4.4) implies that

$$\lim_{h\to 0} \int_{M_\theta'} \Big(\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h} - \dot{s}_\theta\Big)^2 d\mu = \lim_{h\to 0} \int_{M_\theta'} \Big(\frac{\sqrt{f_{\theta+h}} - \sqrt{f_\theta}}{h}\Big)^2 d\mu = 0.$$

Hence, the assertion. □

Proof:: (of Theorem (2.8))
First we note that E is continuous which implies that $\theta \mapsto \sqrt{f_\theta}$ is L^2–continuous. Now the assertion of the Theorem is an immediate consequence of Theorem (4.5) since a continuous function on $\Theta \subseteq \mathbb{R}^d$ is continuously differentiable iff every restriction to a straight line is continuously differentiable (apply Dieudonné, [3], (8.9.1)). □

Proof:: (of Theorem (2.9))
It follows from Lemma (4.1) that the condition is necessary for L^1–differentiability. For the converse we first note that $\theta \mapsto \dot{f}_\theta$ is L^1–continuous. This follows from Lemmas (5.6) and (5.7). Thus (2.10) implies

$$\limsup_{t\to 0} \frac{1}{|t|} \int |f_{\theta+th} - f_\theta| \, d\mu \le \int |\dot{f}_\theta \cdot h| \, d\mu.$$

Since we have

$$\frac{f_{\theta+th} - f_\theta}{t} - \dot{f}_\theta \cdot h \xrightarrow{\mu} 0 \quad \text{if } t \to 0$$

the Lemma of Scheffe gives

$$\lim_{t\to 0} \int \Big| \frac{f_{\theta+th} - f_\theta}{t} - \dot{f}_\theta \cdot h \Big| \, d\mu = 0$$

which proves that all partial L^1–derivatives of $\theta \mapsto f_\theta$ exist and are continuous. Hence the assertion. □

5 Auxiliary lemmas

In this section we collect well–known technical lemmas together with proofs for the reader's convenience.

The first lemma provides an inequality used in the proof of Theorem (2.3).

(5.1) Lemma *Let $a, b, c \in \mathbb{R}$, $a \geq 0,\ b \geq 0$. Then*

$$|a - b - c\sqrt{b}| \leq 2\Big|\sqrt{a} - \sqrt{b} - \frac{c}{2}\Big|\sqrt{b} + (\sqrt{a} - \sqrt{b})^2.$$

Proof: We have

$$\begin{aligned} |a - b - c\sqrt{b}| &= |(\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b}) - c\sqrt{b}| \\ &\leq \Big|(\sqrt{a} - \sqrt{b})\sqrt{a} - \frac{c}{2}\sqrt{b}\Big| + \Big|(\sqrt{a} - \sqrt{b})\sqrt{b} - \frac{c}{2}\sqrt{b}\Big| \\ &= \Big|(\sqrt{a} - \sqrt{b})^2 + (\sqrt{a} - \sqrt{b})\sqrt{b} - \frac{c}{2}\sqrt{b}\Big| + \Big|(\sqrt{a} - \sqrt{b})\sqrt{b} - \frac{c}{2}\sqrt{b}\Big| \\ &\leq 2\Big|\sqrt{a} - \sqrt{b} - \frac{c}{2}\Big|\sqrt{b} + (\sqrt{a} - \sqrt{b})^2. \end{aligned}$$

□

The following lemma prepares a basic fact on the square root of differentiable functions.

(5.2) Lemma *Suppose that f is a continuous and positive function on $[a, b]$. Let $\dot{f}$ be an integrable function on $[a, b]$ such that*

$$f(x_2) - f(x_1) = \int_{x_1}^{x_2} \dot{f}(t)\, dt \quad \text{for every pair } x_1 < x_2 \text{ in } [a, b].$$

Then we have

$$\sqrt{f(x_2)} - \sqrt{f(x_1)} = \int_{x_1}^{x_2} \frac{\dot{f}(t)}{2\sqrt{f(t)}}\, dt \quad \text{for every pair } x_1 < x_2 \text{ in } [a, b].$$

Proof: Since f is bounded away from zero on $[a, b]$ it follows that

$$|\sqrt{f(x_2)} - \sqrt{f(x_1)}| \leq C|f(x_2) - f(x_1)|$$

for some constant $C < \infty$ and every pair $x_1 < x_2$ in $[a, b]$. Since, by assumption, f is an absolutely continuous function on $[a, b]$, its square root $\sqrt{f}$ is an absolutely continuous function on $[a, b]$, too. Moreover, from

$$\sqrt{f(t + h)} - \sqrt{f(t)} = \frac{f(t + h) - f(t)}{(\sqrt{f(t + h)} + \sqrt{f(t)})}$$

it follows that its derivative is

$$(\sqrt{f})' = \frac{\dot{f}}{2\sqrt{f}} \text{ a.e.}$$

Now, from Hewitt–Stromberg, [5], (18.16), the assertion follows. $\square$

The following lemma contains a basic fact on real functions which has been applied by Hájek, [4].

(5.3) LEMMA *Let f be a continuous and nonnegative function on an open interval (α, β). Let $\dot{f}$ be an integrable function on (α, β) such that*

$$f(x_2) - f(x_1) = \int_{x_1}^{x_2} \dot{f}(t)\, dt \quad \textit{for every pair } x_1 < x_2 \textit{ in } (\alpha, \beta).$$

Define

$$\dot{s}(t) := \begin{cases} \dfrac{\dot{f}(t)}{2\sqrt{f(t)}} & \textit{where } f(t) > 0, \\ 0 & \textit{elsewhere.} \end{cases}$$

If $\dot{s}$ is integrable on every compact subinterval of (α, β) then

$$\sqrt{f(x_2)} - \sqrt{f(x_1)} = \int_{x_1}^{x_2} \dot{s}(t)\, dt \quad \textit{for every pair } x_1 < x_2 \textit{ in } (\alpha, \beta).$$

Proof: Let $x_1 < x_2$ be any points in (α, β). Since f is continuous, the interval (x_1, x_2) can be written as

$$(x_1, x_2) = \bigcup_{i=1}^{\infty} (y_i, z_i) \cup M,$$

where (y_i, z_i) are intervals such that $f(x) > 0$ if $x \in (y_i, z_i)$, $f(y_i) = 0$ if $y_i \neq x_1$ and $f(z_i) = 0$ if $z_i \neq x_2$, and where $f = 0$ on M. It is then obvious that

$$\sqrt{f(x_2)} - \sqrt{f(x_1)} = \sum_{i=1}^{\infty} \left(\sqrt{f(z_i)} - \sqrt{f(y_i)} \right).$$

Let us consider a single interval (y, z) where f is positive. If $\epsilon > 0$ then it follows from Lemma (5.2) that

$$\sqrt{f(z-\epsilon)} - \sqrt{f(y+\epsilon)} = \int_{y+\epsilon}^{z-\epsilon} \dot{s}(t)\, dt.$$

By continuity of f it follows that

$$\sqrt{f(z)} - \sqrt{f(y)} = \int_{y}^{z} \dot{s}(t)\, dt.$$

Thus, we obtain

$$\sqrt{f(x_2)} - \sqrt{f(x_1)} = \sum_{i=1}^{\infty} \int_{y_i}^{z_i} \dot{s}(t)\, dt.$$

Since $\dot{s}$ is assumed to be integrable over (x_1, x_2) we may interchange the infinite sum and the integration which gives the assertion. □

Also the arguments used in the following two lemmas are taken from Hájek, [4].

(5.4) LEMMA *Let $\Theta = (\alpha, \beta)$ and assume that $\theta \mapsto ||\dot{f}_\theta||_1$ is continuous. Then there is a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ and*

$$\int_a^b |\dot{f}_t(\omega)|\, dt < \infty \quad \textit{if } a, b \in (\alpha, \beta) \textit{ and } \omega \notin N.$$

Proof: For every $\theta \in (\alpha, \beta)$ there is a $\delta(\theta) > 0$ such that

$$\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} ||\dot{f}_t||_1\, dt < \infty.$$

Hence, for every $\theta \in (\alpha, \beta)$ there is a set $N_\theta \in \mathcal{A}$ such that $\mu(N_\theta) = 0$ and

$$\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} |\dot{f}_t(\omega)|\, dt < \infty \quad \text{if } \omega \notin N_\theta.$$

The family of sets $(\theta - \delta(\theta), \theta + \delta(\theta))$ is a covering of (α, β). There exists a countable subcovering $(\theta_i - \delta(\theta_i), \theta_i + \delta(\theta_i)),\ i \in \mathbb{N}$, of (α, β). We define

$$N := \bigcup_{i=1}^{\infty} N_{\theta_i}$$

and observe that $\mu(N) = 0$. Since $[a, b]$ is covered by finitely many intervals $(\theta_i - \delta(\theta_i), \theta_i + \delta(\theta_i)$, $i = 1, 2, \ldots, M$, it follows that

$$\int_a^b |\dot{f}_t(\omega)|\, dt \leq \sum_{i=1}^{M} \int_{\theta_i-\delta(\theta_i)}^{\theta_i+\delta(\theta_i)} |\dot{f}_t(\omega)|\, dt < \infty \quad \text{if } \omega \notin N.$$

□

(5.5) LEMMA *Let $\Theta = (\alpha, \beta)$ and assume that $\theta \mapsto ||\dot{s}_\theta||_2$ is continuous. Then there is a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ and*

$$\int_a^b |\dot{s}_t(\omega)|\, dt < \infty \quad \textit{if } a, b \in (\alpha, \beta) \textit{ and } \omega \notin N.$$

Proof: For every $\theta \in (\alpha, \beta)$ there is a $\delta(\theta) > 0$ such that

$$\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} ||\dot{s}_t||_2 \, dt < \infty.$$

Hence, for every $\theta \in (\alpha, \beta)$ there is a set $N_\theta \in \mathcal{A}$ such that $\mu(N_\theta) = 0$ and

$$\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} \dot{s}_t(\omega)^2 \, dt < \infty \quad \text{if } \omega \notin N_\theta.$$

Since

$$\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} |\dot{s}_t(\omega)| \, dt \leq \sqrt{2\delta(\theta)} \Big(\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} \dot{s}_t(\omega)^2 \, dt \Big)^{1/2}$$

we obtain

$$\int_{\theta-\delta(\theta)}^{\theta+\delta(\theta)} |\dot{s}_t(\omega)| \, dt < \infty \quad \text{if } \omega \notin N_\theta.$$

The family of sets $(\theta - \delta(\theta), \theta + \delta(\theta))$ is a covering of (α, β). There exists a countable subcovering $(\theta_i - \delta(\theta_i), \theta_i + \delta(\theta_i))$, $i \in \mathbb{N}$, of (α, β). We define

$$N := \bigcup_{i=1}^{\infty} N_{\theta_i}$$

and observe that $\mu(N) = 0$. Since $[a, b]$ is covered by finitely many intervals $(\theta_i - \delta(\theta_i), \theta_i + \delta(\theta_i))$, $i = 1, 2, \ldots, M$, it follows that

$$\int_a^b |\dot{s}_t(\omega)| \, dt \leq \sum_{i=1}^{M} \int_{\theta_i-\delta(\theta_i)}^{\theta_i+\delta(\theta_i)} |\dot{s}_t(\omega)| \, dt < \infty \quad \text{if } \omega \notin N.$$

$\square$

In the following we consider functions $f_\theta \geq 0$, $\dot{f}_\theta$ and $\dot{s}_\theta$ which are related in the following way:

$$\dot{s}(\omega, \theta) := \begin{cases} \dfrac{\dot{f}_\theta(\omega)}{2\sqrt{f(\omega, \theta)}} & \text{where } f(\omega, \theta) > 0, \\ 0 & \text{elsewhere.} \end{cases}$$

and

$$\dot{f}_\theta = 2\dot{s}_\theta \sqrt{f_\theta}$$

(5.6) LEMMA *Suppose that (f_θ) is a μ–continuous family of densities. If $\theta \mapsto \dot{s}_\theta$ is L^2–continuous then $\theta \mapsto \dot{f}_\theta$ is L^1–continuous.*

Proof: We note that

$$\int |\dot f_{\theta_1} - \dot f_{\theta_2}|\,d\mu \le 2\int |\dot s_{\theta_1}\sqrt{f_{\theta_1}} - \dot s_{\theta_2}\sqrt{f_{\theta_2}}|\,d\mu$$

$$\le 2\int |\dot s_{\theta_1} - \dot s_{\theta_2}|\sqrt{f_{\theta_1}}\,d\mu + 2\int |\dot s_{\theta_2}|\,|\sqrt{f_{\theta_1}} - \sqrt{f_{\theta_2}}|\,d\mu$$

$$\le 2\Big(\int (\dot s_{\theta_1} - \dot s_{\theta_2})^2\,d\mu\Big)^{1/2} + 2\Big(\int \dot s_{\theta_2}^2\,d\mu\Big)^{1/2}\Big(\int (\sqrt{f_{\theta_1}} - \sqrt{f_{\theta_2}})^2\,d\mu\Big)^{1/2}.$$

□

(5.7) LEMMA *Suppose that (f_θ) is a μ–continuous family of densities. If $\theta \mapsto \dot f_\theta$ is μ–continuous and $\theta \mapsto ||\dot s_\theta||_2$ is continuous then $\theta \mapsto \dot s_\theta$ is L^2–continuous.*

Proof: By assumption we have

$$\lim_{h\to 0}\int \dot s_{\theta+h}^2\,d\mu = \int \dot s_\theta^2\,d\mu.$$

Let $M_\theta = \{f_\theta > 0\}$. This implies

$$\limsup_{h\to 0}\int_{M_\theta} \dot s_{\theta+h}^2\,d\mu \le \int \dot s_\theta^2\,d\mu = \int_{M_\theta} \dot s_\theta^2\,d\mu. \tag{5.8}$$

Since $\theta \mapsto \dot f_\theta$ and $\theta \mapsto f_\theta$ are μ–continuous it follows that

$$\dot s_{\theta+h} \xrightarrow{\mu} \dot s_\theta \quad \text{on } M_\theta \text{ if } h \to 0.$$

Then the Lemma of Scheffe implies

$$\lim_{h\to 0}\int_{M_\theta} (\dot s_{\theta+h} - \dot s_\theta)^2\,d\mu = 0.$$

Thus we obtain that in (5.8) even equality holds which implies

$$\lim_{h\to 0}\int_{M_\theta'} (\dot s_{\theta+h} - \dot s_\theta)^2\,d\mu = \lim_{h\to 0}\int_{M_\theta'} \dot s_{\theta+h}^2\,d\mu = 0.$$

This proves the assertion. □

References

[1] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins Univ. Press, 1993.

[2] H. Cramer. *Mathematical Methods of Statistics.* Princeton University Press, 1946.

[3] J. Dieudonné. *Foundations of modern analysis.* Academic Press, 1960.

[4] J. Hájek. Local asymptotic minimax and admissibility in estimation. In *Proc. 6th Berkeley Symp. Math. Stat. Prob.*, pages 175–194, 1972.

[5] E. Hewitt and K. Stromberg. *Real and Abstract Analysis.* Springer, Berlin, 1965.

[6] L. LeCam. On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statistics*, 41:802–828, 1970.

[7] L. LeCam. *Asymptotic Methods in Statistical Decision Theory.* Springer, 1986.

[8] L. LeCam and G.L. Yang. *Asymptotics in Statistics.* Springer Series in Statistics. Springer, 1990.

[9] J. Pfanzagl. *Asymptotic expansions for general statistical models*, volume 31 of *Lecture Notes in Statistics.* Springer, 1985.

[10] J. Pfanzagl and W. Wefelmeyer. *Contributions to a general asymptotic statistical theory*, volume 13 of *Lecture Notes in Statistics.* Springer, 1982.

[11] H. Strasser. *Mathematical theory of statistics: Statistical experiments and asymptotic decision theory*, volume 7 of *De Gruyter Studies in Mathematics.* de Gruyter, 1985.

[12] A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, 54:426–482, 1943.

[13] H. Witting. *Mathematische Statistik I.* Teubner, 1985.

Helmut Strasser
Department of Statistics
Vienna University of Economics and Business Administration
Augasse 2–6
A–1090 Vienna
Austria