Ann. Inst. Statist. Math.
29 (1977), Part A, 189-201

CRAMÉR-TYPE CONDITIONS AND QUADRATIC MEAN DIFFERENTIABILITY*

BRUCE LIND AND GEORGE ROUSSAS

(Received Dec. 24, 1973)

Summary

Let $(\Omega, \mathcal{A})$ be a measurable space, let Θ be an open set in R^k, and let $\{P_\theta;\ \theta \in \Theta\}$ be a family of probability measures defined on $\mathcal{A}$. Let μ be a σ-finite measure on $\mathcal{A}$, and assume that $P_\theta \ll \mu$ for each $\theta \in \Theta$. Let us denote a specified version of $dP_\theta/d\mu$ by $f(\omega; \theta)$.

In many large sample problems in statistics, where a study of the log-likelihood is important, it has been convenient to impose conditions on $f(\omega; \theta)$ similar to those used by Cramér [2] to establish the consistency and asymptotic normality of maximum likelihood estimates. These are of a purely analytical nature, involving two or three pointwise derivatives of $\ln f(\omega; \theta)$ with respect to θ. Assumptions of this nature do not have any clear probabilistic or statistical interpretation.

In [10], LeCam introduced the concept of differentially asymptotically normal (DAN) families of distributions. One of the basic properties of such a family is the form of the asymptotic expansion, in the probability sense, of the log-likelihoods. Roussas [14] and LeCam [11] give conditions under which certain Markov Processes, and sequences of independent identically distributed random variables, respectively, form DAN families of distributions. In both of these papers one of the basic assumptions is the differentiability in quadratic mean of a certain random function. This seems to be a more appealing type of assumption because of its probabilistic nature.

In this paper, we shall prove a theorem involving differentiability in quadratic mean of random functions. This is done in Section 2. Then, by confining attention to the special case when the random function is that considered by LeCam and Roussas, we will be able to show that the standard conditions of Cramér type are actually stronger than the conditions of LeCam and Roussas in that they imply the existence of the necessary quadratic mean derivative. The relevant discussion is found in Section 3.

* This research was supported by the National Science Foundation, Grant GP-20036.

1. Introduction

Before stating the theorems, we must introduce the notation to be used. Let $(\Omega, \mathcal{A}, P)$ be a probability space and denote by

$$L_2(\Omega, P)=L_2(\Omega)=\{\text{random variables (r.v.'s), } X, \text{ on } (\Omega, \mathcal{A}, P);\ \mathcal{E}X^2<\infty\} .$$

For $X, Y \in L_2(\Omega)$, define the inner product $\langle X, Y\rangle$ as follows

$$\langle X, Y\rangle=\mathcal{E}(XY) .$$

Denote by $\|\cdot\|_2$ the L_2-norm induced by the inner product $\langle\cdot, \cdot\rangle$; i.e., for $X \in L_2(\Omega)$,

$$\| X\|_2=(\langle X, X\rangle)^{1/2} .$$

Next, let Θ be a k-dimensional, open subset of R^k and let $g(\theta)=g(\theta_1, \theta_2, \cdots, \theta_k)$, $\theta=(\theta_1, \theta_2, \cdots, \theta_k)'$ be a random function on $(\Omega, \mathcal{A}, P)$, where the random element ω is omitted from the notation for the sake of simplicity.

DEFINITION 1.1. The random function $g(\theta)$ is said to be differentiable in quadratic mean (q.m.) at θ when P is employed if there exists a k-dimensional vector of random functions, $\dot{g}(\theta)$, such that

$$|h|^{-1}\| g(\theta+h)-g(\theta)-h'\dot{g}(\theta)\|_2\rightarrow 0$$

as $0\neq|h|\rightarrow 0$, where $|\cdot|$ denotes the usual Euclidean norm of the vector h; $\dot{g}(\theta)$ is the q.m. derivative of $g(\theta)$ at θ. Here, "$'$" denotes transpose and $h'\dot{g}(\theta)$ is the inner product of the indicated vectors.

2. Statements and proofs of some theorems on quadratic mean derivatives

The first theorem is essentially a special case of a result of LeCam [12]. In it, we consider the case where $k=1$; i.e., θ is a real parameter.

THEOREM 2.1. *Assume the notation of Section* 1. *Let* $g(\omega, \theta)$ *be a random function defined on* $\Omega\times\Theta$ *which is jointly measurable in* ω *and* θ, *and such that*

(i) $(\partial g(\omega, \theta)/\partial\theta)|_{\theta=\theta^*}=g_2(\omega, \theta^*) \in L_2(\Omega, P_{\theta_0})$ *for all* θ^* *in some neighborhood of* $\theta_0 \in \Theta$.

Suppose furthermore that

(ii) $g_2(\omega, \theta^*)$ *is finite, except possibly at countably many points, in a neighborhood of* θ_0, *a.s.* $[P_{\theta_0}]$.

Let $h^2(\theta^*)=\int_\Omega g_2^2(\omega,\theta^*)P_{\theta_0}(d\omega)$, *and assume that*

(iii) $\lim t^{-1}\int_{\theta_0}^{\theta_0+t}|h(u)-h(\theta_0)|du=0$ *as* $(0\neq)\, t\to 0$.

Then, $g_2(\omega,\theta_0)$ *is the q.m.* ($[P_{\theta_0}]$) *derivative of* $g(\omega,\theta)$ *at* $\theta=\theta_0$.

PROOF. It follows from the above hypotheses (i) and (ii) and Theorem 264 of Kestelman [8], that for every θ_1, θ_2 in a neighborhood of θ_0 and almost all ($[P_{\theta_0}]$) $\omega\in\Omega$, we have

$$g(\omega,\theta_2)-g(\omega,\theta_1)=\int_{\theta_1}^{\theta_2}g_2(\omega,u)du\ . \tag{2.1}$$

Next,

$$\begin{aligned}
&\|t^{-1}[g(\omega,\theta_0+t)-g(\omega,\theta_0)]\|_2^2\\
&=\int_\Omega t^{-2}[g(\omega,\theta_0+t)-g(\omega,\theta_0)]^2P_{\theta_0}(d\omega)\\
&=\int_\Omega\left|t^{-1}\int_{\theta_0}^{\theta_0+t}g_2(\omega,u)du\right|^2P_{\theta_0}(d\omega) && (2.2)\\
&\leq t^{-2}\int_\Omega\left(\int_{\theta_0}^{\theta_0+t}|g_2(\omega,u)|du\int_{\theta_0}^{\theta_0+t}|g_2(\omega,v)|dv\right)P_{\theta_0}(d\omega)\\
&=t^{-2}\int_{\theta_0}^{\theta_0+t}\int_{\theta_0}^{\theta_0+t}\left[\int_\Omega|g_2(\omega,u)||g_2(\omega,v)|P_{\theta_0}(d\omega)\right]dudv && (2.3)\\
&\leq t^{-2}\int_{\theta_0}^{\theta_0+t}\int_{\theta_0}^{\theta_0+t}h(u)h(v)dudv && (2.4)\\
&=\left|t^{-1}\int_{\theta}^{\theta_0+t}h(u)du\right|^2 && (2.5)\\
&=\left|t^{-1}\int_{\theta_0}^{\theta_0+t}h(u)du-t^{-1}h(\theta_0)t+h(\theta_0)\right|^2\\
&=\left|h(\theta_0)+t^{-1}\int_{\theta_0}^{\theta_0+t}[h(u)-h(\theta_0)]du\right|^2,
\end{aligned}$$

where (2.2) follows from (2.1), (2.3) results from a double application of Fubini's theorem, (2.4) is obtained by Schwarz inequality and (2.5) by Fubini's theorem. That is,

$$\|t^{-1}[g(\omega,\theta_0+t)-g(\omega,\theta_0)]\|_2^2\leq\left|h(\theta_0)+t^{-1}\int_{\theta_0}^{\theta_0+t}[h(u)-h(\theta_0)]du\right|^2$$

and hence, as $(0\neq)\, t\to 0$, we obtain by means of (iii)

$$\limsup\|t^{-1}[g(\omega,\theta_0+t)-g(\omega,\theta_0)]\|_2^2\leq h^2(\theta_0)\ .$$

On the other hand, by Fatou's lemma, one has

$$h^2(\theta_0)\leq\liminf\|t^{-1}[g(\omega,\theta_0+t)-g(\omega,\theta_0)]\|_2^2 \qquad \text{as } (0\neq)\, t\to 0\ .$$

Combining these last two results, we get

$$\|t^{-1}[g(\omega, \theta_0+t)-g(\omega, \theta_0)]\|_2^2 \to h^2(\theta_0) \qquad \text{as } (0\neq)t\to 0 .$$

Since also

$$\{t^{-1}[g(\omega, \theta_0+t)-g(\omega, \theta_0)]\}^2 \xrightarrow{P_{\theta_0}} g_2^2(\omega, \theta_0) \qquad \text{as } (0\neq)\, t\to 0 ,$$

one has

$$t^{-1}[g(\omega, \theta_0+t)-g(\omega, \theta_0)] \to g_2(\omega, \theta_0) \qquad \text{in q.m. } ([P_{\theta_0}]) \text{ as } (0\neq)t\to 0 ,$$

which establishes the desired result.

All of the hypotheses of the theorem can be readily checked except, perhaps, for (iii). The following lemma will be helpful in formulating an alternative condition.

LEMMA 2.1. *Let* $q: R\to R$, *and let* x_0 *be a continuity point of* q. *Let* $|q|$ *be Lebesgue integrable in a neighborhood of* x_0. *Then, as* $(0\neq)$ $t\to 0$,

$$\lim t^{-1} \int_{x_0}^{x_0+t} |q(u)-q(x_0)|\, du=0 .$$

PROOF. Since q is continuous at x_0, for a given $\varepsilon>0$, there exists a $\delta>0$ such that

$$|q(u)-q(x_0)|<\varepsilon \qquad \text{if } |u-x_0|<\delta .$$

If we choose $|t|<\delta$, we have

$$t^{-1} \int_{x_0}^{x_0+t} |q(u)-q(x_0)|\, du<\varepsilon$$

and the lemma is proven.

Using the above lemma we see that (iii) is implied by
(iii)′ $h(\theta)$ is locally integrable and continuous at θ_0.

To close this section we state, without proof, a theorem which has been derived in an earlier paper (Lind and Roussas [13]). It deals with the case $k>1$, and will be referred to in Section 3.

THEOREM 2.2. *For each* $\theta\in\Theta$, *we assume that the partial derivatives in q.m.* ($[P_\theta]$) *of the random function* $g(\theta)$ *exist and are continuous (in* θ*) in* L_2*-norm,* $\|\cdot\|_2$. *Then* $\dot{g}(\theta)$, *the derivative in q.m. of* $g(\theta)$, *exists and is equal to the vector of partial derivatives in q.m. That is, for each* $\theta\in\Theta$ *and* $h\neq 0$,

$$\left\| |h|^{-1}\left[g(\theta+h)-g(\theta)-\sum_{j=1}^{k} h_j\dot{g}_j(\theta)\right]\right\|_2 \to 0$$

as $|h|\to 0$, *where* $\dot{g}_j(\theta)$ *denotes the partial q.m. derivative of* $g(\theta)$ *at* θ.

This theorem changes the problem from a search for a vector of functions and the taking of limits as a vector variable approaches zero in norm to one of finding the partial derivatives in q.m., which involves only a single function and a real argument, and then checking to see that they are continuous in $\|\cdot\|_2$-norm. Theorem 2.1 may serve as a way of finding the partial q.m. derivatives.

Further use of these theorems will be demonstrated in a forthcoming paper which will be devoted to the calculation of some asymptotically optimal tests for certain failure distributions.

3. Comparison of Cramér type conditions with the conditions of LeCam and Roussas

In LeCam [12], a set of conditions of the Cramér type is shown to imply a certain differentiability in quadratic mean assumption of LeCam and Roussas. This set of conditions is essentially minimal and the proof is quite involved and makes use of some techniques of functional analysis. In this section, it will be shown that, if one instead chooses the stronger conditions usually made, then the conclusion can be arrived at by some simple arguments.

To facilitate the comparison of the two types of conditions for the remainder of this section, let $(\Omega, \mathcal{A})$ be $(R, \mathcal{B})$ the Borel real line, let Θ be an open subset of R^k, and let $\{P_\theta;\ \theta \in \Theta\}$ be a family of probability measures defined on $\mathcal{B}$. Let μ be a σ-finite measure on $\mathcal{B}$, and assume that $P_\theta \ll \mu$ for each $\theta \in \Theta$. Let us denote a specified version of $dP_\theta/d\mu$ by $f(x;\theta)$. Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed random variables defined on $(R, \mathcal{B})$. Then the assumptions of Roussas [14] are the following:

(A1) The set on which $f(\cdot;\theta)$ is positive is independent of θ.

Set $\phi_1(\theta;\theta^*) = [f(X_1;\ \theta^*)/f(X_1;\theta)]^{1/2}$. Then

(A2) (i) For each $\theta \in \Theta$, the random function $\phi_1(\theta;\theta^*)$ is differentiable in q.m. with respect to θ^* at (θ, θ) when P_θ is employed.

Let $\dot{\phi}_1(\theta)$ be the q.m. derivative of $\phi_1(\theta;\theta^*)$ with respect to θ^* at (θ, θ). Then

(ii) $\dot{\phi}_1(\theta)$ is $X_1^{-1}(\mathcal{B}) \times C$-measurable, where $\mathcal{B}$ is the Borel σ-field in R' and C is the σ-field of Borel subsets of Θ.

(iii) For every $\theta \in \Theta$, $\Gamma(\theta) = 4\mathcal{E}_\theta[\dot{\phi}_1(\theta)\dot{\phi}_1'(\theta)]$ is positive definite.

We now state a set of conditions of Cramér type. These are found in Davidson and Lever [3].

(B1) Same as (A1) above.

(B2) For almost all $([\mu])x \in R$ and for all $\theta \in \Theta$,

$$\partial \ln f/\partial\theta_r\,, \quad \partial^2 \ln f/\partial\theta_r\partial\theta_s \quad \text{and} \quad \partial^3 \ln f/\partial\theta_r\partial\theta_s\partial\theta_t$$

exist for $r, s, t=1,\cdots, k$.

(B3) For almost all $([\mu])x \in R$ and for every $\theta \in \Theta$,

$$|\partial f/\partial\theta_r|<F_r(x) \quad \text{and} \quad |\partial^2 f/\partial\theta_r\partial\theta_s|<F_{rs}(x) ,$$

where $F_r(x)$ and $F_{rs}(x)$ are integrable over R, $r, s=1,\cdots, k$.

(B4) For every $\theta \in \Theta$, the matrix $I(\theta)=(I_{rs}(\theta))$ with

$$I_{rs}(\theta)=\mathcal{E}_\theta\left(\frac{\partial \ln f}{\partial\theta_r}\Big|_\theta \frac{\partial \ln f}{\partial\theta_s}\Big|_\theta\right)$$

is positive definite with finite determinant.

(B5) For almost all $([\mu])x \in R$ and for all $\theta \in \Theta$,

$$\left|\frac{\partial^3 \ln f}{\partial\theta_r\partial\theta_s\partial\theta_t}\Big|_\theta\right|<H_{rst}(x) ,$$

where there exists a positive real number M such that

$$\mathcal{E}_\theta[H_{rst}(X_1)]<M<\infty$$

for all $\theta \in \Theta$ and $r, s, t=1,\cdots, k$.

These assumptions are simply a multi-parameter extension of those given by Cramér [2].

When one wants to derive more delicate results for a family of distributions it is convenient to add some further hypotheses on the density. One purpose of these further restrictions is to insure the continuity (componentwise) of $I(\theta)$. (See Kaufman [7], Weiss and Wolfowitz [20] and LeCam [9] for papers referring to estimation, and Wald [18], [19], Davidson and Lever [3] and LeCam [9] for papers referring to tests of hypotheses.) Since the conditions (A), along with the assumption that $\Gamma(\theta)$ is continuous, are sufficient to provide results of the same nature, it is not unreasonable to incorporate a condition which implies the continuity of $I(\theta)$ into (B) above. The one we choose to use is from Davidson and Lever [3] (see also Wald [18], [19]).

(B6) There exist positive real numbers ν and T such that whenever

$$\|\theta''-\theta'\| \overset{\text{df}}{=} \sum_{r=1}^{k} |\theta''_r-\theta'_r|<\nu , \qquad \theta', \theta'' \in \Theta ,$$

$$\mathcal{E}_{\theta'}\left[\left(\frac{\partial^2 \ln f}{\partial\theta_r\partial\theta_s}\Big|_{\theta''}\right)^2\right]<T<\infty \qquad \text{for } r, s=1,\cdots, k .$$

Since the proof that (B6) does, in fact, imply that $I(\theta)$ is continuous is quite interesting, and not included in the Davidson and Lever paper, we include it here. In other words, we establish the following result

THEOREM 3.1. *Under assumptions* (B1)–(B6)

$$I(\theta)=\left(\mathcal{E}_\theta\left(\frac{\partial \ln f}{\partial \theta_r}\bigg|_\theta \frac{\partial \ln f}{\partial \theta_s}\bigg|_\theta\right)\right)$$

is continuous in the sense that each component is a continuous function of θ.

PROOF. It is well known that (B1)-(B5) imply that

$$I(\theta)=\left(-\mathcal{E}_\theta\left(\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_\theta\right)\right).$$

Therefore,

$$\begin{aligned}
&I_{rs}(\theta)-I_{rs}(\theta_0)\\
&\quad=-\int_R \frac{\partial^2 \ln f(x;\theta)}{\partial \theta_r \partial \theta_s}\bigg|_\theta f(x;\theta)d\mu(x)+\int_R \frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0} f(x;\theta_0)d\mu(x)\\
&\quad=\int_R \left[\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0} f(x;\theta_0)-\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_\theta f(x;\theta)\right]d\mu(x)\,.
\end{aligned}$$

By Taylor's theorem, we have for almost all $([\mu])x \in R$,

$$\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_\theta=\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0}+\sum_{t=1}^{k}(\theta_t-\theta_{0t})\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}\bigg|_{\theta^*},$$

where θ^* lines on the line segment joining θ and θ_0. From this, we have

$$\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_\theta f(x;\theta)=\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0} f(x;\theta)+\sum_{t=1}^{k}(\theta_t-\theta_{0t})\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}\bigg|_{\theta^*} f(x;\theta)$$

which implies

$$\begin{aligned}
&\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0} f(x;\theta_0)-\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_\theta f(x;\theta)\\
&\quad=\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0}[f(x;\theta_0)-f(x;\theta)]-\sum_{t=1}^{k}(\theta_t-\theta_{0t})\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}\bigg|_{\theta^*} f(x;\theta)\,.
\end{aligned}$$

Thus,

$$\begin{aligned}
&|I_{r,s}(\theta)-I_{r,s}(\theta_0)|\\
&\quad\leqq\left|\int_R \frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0}[f(x;\theta_0)-f(x;\theta)]d\mu(x)\right|\\
&\qquad+\left|\sum_{t=1}^{k}(\theta_t-\theta_{0t})\int_R \frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}\bigg|_{\theta^*} f(x;\theta)d\mu(x)\right|\\
&\quad\leqq\int_R \left|\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}\bigg|_{\theta_0}\right| ||f(x;\theta_0)-f(x;\theta)|^{1/2}|f(x;\theta_0)-f(x;\theta)|^{1/2}d\mu(x)|\\
&\qquad+\sum_{t=1}^{k}|\theta_t-\theta_{0t}|\int_R \left|\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}\bigg|_{\theta^*}\right| f(x;\theta)d\mu(x)\,.
\end{aligned}$$

By the Cauchy-Schwarz inequality this is less than or equal to

$$\left\{\int_R \left|\frac{\partial^2 \ln f}{\partial\theta_r\partial\theta_s}\bigg|_{\theta_0}\right|^2 |f(x;\theta_0)-f(x;\theta)|\,d\mu(x)\right\}^{1/2}$$
$$\cdot\left\{\int_R |f(x;\theta_0)-f(x;\theta)|\,d\mu(x)\right\}^{1/2}$$
$$+\sum_{t=1}^{k}|\theta_t-\theta_{0t}|\int_R\left|\frac{\partial^3 \ln f}{\partial\theta_r\partial\theta_s\partial\theta_t}\bigg|_{\theta^*}\right| f(x;\theta)\,d\mu(x)\;.$$

By assumption (B5), the last sum above is bounded by $M\|\theta-\theta_0\|$. The first term is bounded by

$$\left\{\int_R \left|\frac{\partial^2 \ln f}{\partial\theta_r\partial\theta_s}\bigg|_{\theta_0}\right|^2 f(x;\theta_0)\,d\mu(x)+\int_R \left|\frac{\partial^2 \ln f}{\partial\theta_r\partial\theta_s}\bigg|_{\theta_0}\right|^2 f(x;\theta)\,d\mu(x)\right\}^{1/2}$$
$$\cdot\left\{\sum_{r=1}^{k}|\theta_r-\theta_{0r}|\int_R\left|\frac{\partial f}{\partial\theta_r}\bigg|_{\theta^{**}}\right|d\mu(x)\right\}^{1/2}$$

where θ^{**} lies on the line segment joining θ and θ_0. By (B6), the first factor above is less than $(2T)^{1/2}$ for $\|\theta-\theta_0\|$ sufficiently small ($<\nu$). By (B3), the second factor is less than $(K\|\theta-\theta_0\|)^{1/2}$, where $K=\int_R F_r(x)d\mu(x)$. Therefore $|I_{r,s}(\theta)-I_{r,s}(\theta_0)|\leqq M\|\theta-\theta_0\|+(2Tk\|\theta-\theta_0\|)^{1/2}$, provided $\|\theta-\theta_0\|<\nu$. From this it follows that $\lim|I_{r,s}(\theta)-I_{r,s}(\theta_0)|=0$, as $\theta\to\theta_0$, as was to be seen.

Let us now assume that $k=1$, and prove the following theorem.

THEOREM 3.2. *In the notation above, assumptions* (B1)–(B6) *are stronger than assumptions* (A1)–(A2) *in the sense that any density which satisfies* (B1)–(B6) *also satisfies* (A1)–(A2).

Before proceeding with the proof of Theorem 3.2, we specialize Theorem 2.1 to the case where the random function is ϕ_1. (In the notation below (i), (ii), and (iii)′ correspond to (i), (ii), and (iii)′ as used in Theorem 2.1 and the discussion following it.)

Recall that

$$\phi_1(x;\theta_0,\theta^*)=[f(x;\theta_0)]^{-1/2}[f(x;\theta^*)]^{1/2}\;.$$

Therefore,

$$\frac{\partial\phi_1(x;\theta_0,\theta^*)}{\partial\theta^*}\bigg|_{\theta^*=\bar\theta}=\frac{1}{2}\,\frac{\partial f(x;\theta^*)}{\partial\theta^*}\bigg|_{\bar\theta}[f(x;\theta_0)f(x;\bar\theta)]^{-1/2}\;.$$

Thus, we want to have

(i) $$\frac{1}{4}\int_R\frac{[(\partial f(x;\theta)/\partial\theta)|_{\theta'}]^2}{f(x;\theta')f(x;\theta_0)}f(x;\theta_0)\,d\mu(x)$$

$$=\frac{1}{4}\int_R\left[\frac{\partial f(x;\theta)}{\partial\theta}\Big|_{\theta'}\Big/f(x;\theta')\right]^2 f(x;\theta')d\mu(x)$$

$$=\frac{1}{4}I(\theta') \qquad \text{finite for } \theta' \text{ in a neighborhood of } \theta_0;$$

(ii) For x (almost all $[\mu]$) we want

$$\frac{1}{2}[f(x;\theta_0)]^{-1/2}\frac{\partial f(x;\theta)}{\partial\theta}\Big|_{\theta'}\Big/[f(x;\theta')]^{1/2}$$

to be finite as a function of θ' (except for possibly countably many values) in a neighborhood of θ_0;

(iii)′ $I(\theta)$ is a continuous function of θ in a neighborhood of θ_0 (since this in addition to (i) above implies the local integrability of $h(\theta) = (1/2)[I(\theta)]^{1/2}$).

In proving Theorem 3.2 let us note that verification of (i), (ii), (iii)′ above for each $\theta_0 \in \Theta$ is sufficient to establish the following

LEMMA 3.1. *Conditions* (B1)–(B6) *imply that* (A2)-(i) *is satisfied, with*

$$\dot{\phi}_1(x;\theta)=\frac{\partial}{\partial\theta^*}\phi_1(x;\theta,\theta^*)|_{\theta^*=\theta}=\frac{1}{2}\frac{\partial \ln f(x;\theta^*)}{\partial\theta^*}\Big|_{\theta^*=\theta}$$

for each $\theta\in\Theta$.

PROOF. Clearly, (i) is satisfied because of (B4). Let us consider condition (ii). Now,

$$\frac{\partial\phi_1(x;\theta_0,\theta^*)}{\partial\theta^*}\Big|_{\theta^*=\theta'}=\frac{1}{2}[f(x;\theta_0)]^{-1/2}\left[\frac{\partial f(x;\theta^*)}{\partial\theta^*}\Big|_{\theta^*=\theta'}\right][f(x;\theta')]^{-1/2}.$$

By (B2) $(\partial \ln f)/\partial\theta$ exists and it follows that both $[(\partial \ln f(x;\theta)/\partial\theta)|_{\theta'}]$ and $[f(x;\theta')]^{1/2}$ are finite functions of θ' and hence (ii) is satisfied. For (iii)′ it suffices to show that $I(\theta)$ is a continuous function of θ. This has been done in Theorem 3.1.

PROOF OF THEOREM 3.2. To complete the proof of Theorem 3.2 we observe that (B1) and (A1) are identical. It only remains to show that (A2)-(ii) and (A2)-(iii) are implied by conditions (B). From Lemma 3.1 we have $\dot{\phi}_1(\theta)=(1/2)((\partial \ln f(x;\theta))/\partial\theta)$, and thus (A2)-(ii) and (A2)-(iii) follow from (B2)–(B4). This completes the proof of Theorem 3.2.

We thus know that for $k=1$, conditions (B) imply conditions (A). The above result also shows that the partial derivatives in quadratic mean are given by the pointwise partial derivatives under conditions (B). If we could also show that conditions (B) imply the continuity in

L_2-norm of these partial derivatives we would have the existence of $\dot{\phi}_1$ in the case $k>1$ upon application of Theorem 2.2.

What we mean by continuity in L_2-norm in this case is the following: Since we have already shown that

$$\dot{\phi}_{1,r}(\theta_0)=\frac{1}{2}\left(\frac{\partial f(x;\theta)}{\partial\theta_r}\bigg|_{\theta_0}\right)\Big/[f(x;\theta_0)] ,$$

where $\dot{\phi}_{1,r}(\theta_0)$ denotes the partial q.m. derivative of $\phi(x;\theta_0,\theta^*)$ with respect to θ_r^* at $\theta^*=\theta_0$, what we want to show is that

$$\lim_{\theta^*\to\theta_0}\frac{1}{4}\int_R\left[\left(\frac{\partial f(x;\theta)}{\partial\theta_r}\bigg|_{\theta^*}\right)\Big/f(x;\theta^*)-\left(\frac{\partial f(x;\theta)}{\partial\theta_r}\bigg|_{\theta_0}\right)\Big/f(x;\theta_0)\right]^2 \cdot f(x;\theta_0)d\mu(x)$$

equals zero. This is equivalent to

$$\lim_{\theta^*\to\theta_0}\int_R\left[\frac{\partial\ln f(x;\theta)}{\partial\theta_r}\bigg|_{\theta^*}-\frac{\partial\ln f(x;\theta)}{\partial\theta_r}\bigg|_{\theta_0}\right]^2 f(x;\theta_0)d\mu(x)=0 .$$

Since $(\partial^2\ln f)/\partial\theta_r\partial\theta_s$ exists $s=1,2,\cdots,k$, we have

$$\frac{\partial\ln f}{\partial\theta_r}\bigg|_{\theta^*}-\frac{\partial\ln f}{\partial\theta_r}\bigg|_{\theta_0}\to 0 \qquad \text{as } \theta^*\to\theta_0 .$$

Thus if we can show

$$\lim_{\theta^*\to\theta_0}\int_R\left[\frac{\partial\ln f}{\partial\theta_r}\bigg|_{\theta^*}\right]^2 f(x;\theta_0)d\mu(x)=\int_R\left[\frac{\partial\ln f}{\partial\theta_r}\bigg|_{\theta_0}\right]^2 f(x;\theta_0)d\mu(x) ,$$

the proof will be complete by Vitali's theorem. (B6) is not quite strong enough to readily give us the desired result. What is needed is a requirement such as is found in Bahadur [1], Roussas [15], or Schmetterer [17].

(B6)′ For any given θ_0 in Θ, and $r=1,\cdots,k$, there exists a neighborhood, N_{θ_0}, of θ_0 and a $\mathcal{B}$-measurable function $M_r(x)$ such that

$$\left|\frac{\partial\ln f}{\partial\theta_r}\right|\leq M_r(x) \qquad \text{for all } x\in R$$

and all $\theta\in N_{\theta_0}$ and such that

$$\mathcal{E}_{\theta_0}M_r^2<\infty , \qquad r=1,\cdots,k .$$

With this condition it is obvious from the dominated convergence theorem that

$$\int_R\left[\frac{\partial\ln f}{\partial\theta_r}\bigg|_{\theta^*}\right]^2 f(x;\theta_0)d\mu(x)$$

is a continuous function of θ^* in a neighborhood of θ_0.

4. Closing remarks and an example

The results presented in this paper justify earlier assertions that the assumptions (A) are actually weaker than those commonly occurring in the literature and not just of a different nature. This shows that the results of Roussas [14], [16] and Johnson and Roussas [4], [5], [6] extend earlier results in two ways. The regularity conditions are weaker and the assumption of independence is also dropped. Since there are examples in which conditions (A) are satisfied when conditions (B) are not, the results actually have wider application as well.

Example. As an example of a density which is, clearly, non-regular with respect to the usual Cramér type conditions, but is regular with respect to the (A) conditions, consider

$$f(x;\theta)=\frac{1}{2\Gamma(1+1/\lambda)}\exp\{-|x-\theta|^{\lambda}\}\,,\qquad 1/2<\lambda\leqq 1\,. \tag{4.1}$$

A discussion of the case $\lambda=1$, the Laplace distribution, is found in Johnson and Roussas [4]. We include here a brief discussion for $1/2<\lambda<1$.

For the density (4.1), we have

$$\phi_1(\theta,\theta^*)=\exp\left\{-\frac{1}{2}|X_1-\theta^*|^{\lambda}+\frac{1}{2}|X_1-\theta|^{\lambda}\right\}\,.$$

Now let

$$g_1(\theta)=\begin{cases}-\frac{1}{2}\lambda(\theta-X_1) & X_1<\theta\\ 0 & X_1=\theta\\ \frac{1}{2}\lambda(X_1-\theta) & X_1>\theta\,.\end{cases}$$

It is clear that

$$h^{-1}[\phi_1(\theta,\theta+h)-1]\to g_1(\theta)\qquad\text{in } P_\theta\text{-probability}$$

as $(0\neq)\,h\to 0$. Next,

$$\mathcal{E}_\theta\{h^{-1}[\phi_1(\theta,\theta+h)-1]\}^2=2h^{-2}[1-\mathcal{E}_\theta\phi_1(\theta,\theta+h)]\,.$$

Thus, if we can show that

$$2h^{-2}[1-\mathcal{E}_\theta\phi_1(\theta,\theta+h)]\to\mathcal{E}_\theta[g_1(\theta)]^2\qquad\text{as } (0\neq)\,h\to 0\,, \tag{4.2}$$

we will have verified (A2)-(i), and the remaining (A) conditions are obviously satisfied for $\dot{\phi}_1(\theta)=g_1(\theta)$.

A straightforward computation gives

$$\mathcal{E}_\theta[g_1(\theta)]^2=\frac{\lambda\Gamma(2-1/\lambda)}{4\Gamma(1+1/\lambda)}\ .$$

To verify (4.2), we first form the expression $2h^{-2}[1-\mathcal{E}_\theta\phi_1(\theta,\theta+h)]$. This is equal to (after some simplification)

$$2h^{-2}\Big[1-\frac{1}{\Gamma(1+1/\lambda)}\int_0^\infty \exp\Big\{-\frac{1}{2}(y+h)^\lambda-\frac{1}{2}y^\lambda\Big\}dy$$
$$-\frac{1}{2\Gamma(1+1/\lambda)}\int_0^h \exp\Big\{-\frac{1}{2}y^\lambda-\frac{1}{2}(h-y)^\lambda\Big\}dy\ .$$

Evaluation of the limit, as $(0\neq)\,h$ approaches zero, can be accomplished by two successive applications of L'Hospital's rule (which is easily justifiable by the Dominated convergence theorem, differentiation of the second integral in the expression immediately above following by Leibowitz's rule). The result of these operations yields $(\lambda\Gamma(2-1/\lambda))/(4\Gamma(1+1/\lambda))$ and the example is completed.

UNIVERSITY OF WISCONSIN

REFERENCES

[1] Bahadur, R. R. (1964). On Fisher's bound for asymptotic variances, *Ann. Math. Statist.*, **35**, 1545-1552.

[2] Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton.

[3] Davidson, R. R. and Lever, W. E. (1970). The limiting distribution of the likelihood ratio statistic under a class of local alternatives, *Sankhya* Ser. A, **32**, 209-224.

[4] Johnson, R. A. and Roussas, G. G. (1969). Asymptotically most powerful tests in Markov processes, *Ann. Math. Statist.*, **40**, 1207-1215.

[5] Johnson, R. A. and Roussas, G. G. (1970). Asymptotically optimal tests in Markov processes, *Ann. Math. Statist.*, **41**, 918-938.

[6] Johnson, R. A. and Roussas, G. G. (1971). Applications of contiguity to multiparameter hypothesis testing, *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, 195-226.

[7] Kaufman, S. (1966). Asymptotic efficiency of the maximum likelihood estimator, *Ann. Inst. Statist. Math.*, 18, 155-178.

[8] Kestelman, H. (1960). *Modern Theories of Integration*, Dover Publications, Inc. New York.

[9] LeCam, L. (1956). On the asymptotic theory of estimation and testing hypotheses, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, 129-156.

[10] LeCam, L. (1960). Locally asymptotically normal families of distributions, *Univ. Calif. Pub. Statist.*, **3**, 37-98.

[11] LeCam, L. (1966). Likelihood functions for large numbers of independent observations, *Research Papers in Statistics* (F. N. David, editor), Wiley, New York.

[12] LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates, *Ann. Math. Statist.*, **41**, 802-828.

[13] Lind, R. B. and Roussas, G. G. (1970). Multiparameter differentiation in quadratic mean and asymptotically optimal tests for some failure distributions, *Technical Report* No. 227, University of Wisconsin, Madison.

[14] Roussas, G. G. (1965). Asymptotic inference in Markov processes, *Ann. Math. Statist.*, **36**, 978-992.

[15] Roussas, G. G. (1968). Asymptotic normality of the maximum likelihood estimate in Markov processes, *Metrika*, **14**, 62-70.

[16] Roussas, G. G. (1968). Some applications of the asymptotic distribution of likelihood functions to the asymptotic efficiency of estimates, *Zeit Wahrscheinlichkeitsth.*, **10**, 252-260.

[17] Schmetterer, L. (1966). On the asymptotic efficiency of estimates, *Research Papers in Statistics* (F. N. David, editor), Wiley, New York.

[18] Wald, A. (1941). Asymptotically most powerful tests of statistical hypotheses, *Ann. Math. Statist.*, **12**, 1-19.

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

[20] Weiss, L. and Wolfowitz, J. (1966). Generalized maximum likelihood estimators, *Theory Prob. Appl.*, **11**, 58-81.