The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 802–828

ON THE ASSUMPTIONS USED TO PROVE ASYMPTOTIC NORMALITY OF MAXIMUM LIKELIHOOD ESTIMATES[1]

BY L. LECAM

University of California, Berkeley

1. Introduction. Let Θ be a subset of the real line. For each $\theta \in \Theta$ let p_θ be a probability measure on a certain σ-field $\mathfrak{A}$ carried by a set $\mathfrak{X}$. Let $\{\mathfrak{X}^n, \mathfrak{A}^n\}$ be the cartesian product of n copies of $\{\mathfrak{X}, \mathfrak{A}\}$. Let $P_{\theta,n}$ be the measure product of n copies of p_θ.

Several statistical problems lead to the study of the behavior of the functions $\theta \rightleftarrows P_{\theta,n}$ as n tends to infinity. In particular, the statements concerning maximum likelihood estimates found in Cramér [1] can be considered statements about the local behavior of the logarithm of the likelihood function

$$\Lambda_n(t, \theta) = \log \frac{dP_{t,n}}{dP_{\theta,n}}$$

when n increases indefinitely.

A similar assertion can be made about the deeper results of Wald [11] concerning the asymptotic sufficiency of the maximum likelihood estimates and the asymptotic normality of the family of measures. (Wald gives convergence results uniformly in θ instead of locally; however, the bulk of the argumentation is local.)

The regularity conditions used by Cramér or Wald or other authors, for instance Doob [2], Dugué [3], Wilks [12] always involve the existence of two or three derivatives of the function $t \rightleftarrows dp_t/dp_\theta$ and additional uniform integrability restrictions. These regularity restrictions do not have by themselves any direct or obvious statistical relevance or interpretation. Their role is to permit the proof of the desired theorems.

It has long been realized that the assumptions in question are somewhat too stringent for this purpose and that in fact the asymptotic normality derived from them ought to be a consequence of assumptions involving only first derivatives.

Even if one is not particularly interested in the maximum economy of assumptions one cannot escape practical statistical problems in which apparently "slight" violations of the assumptions occur. For instance the derivatives fail to exist at one point x which may depend on θ, or the distributions may not be mutually absolutely continuous or a variety of other difficulties may occur. The existing literature is rather unclear about what may happen in these circumstances. Note also that since the conditions are imposed upon probability densities they may be satisfied for one choice of such densities but not for certain other choices.

Received September 11, 1969.

[1] This work was completed with the partial support of U.S. Army Grant DA AROD-31-124 G-816. This paper is submitted in partial fulfillment of the promotion requirement of the University of California, Berkeley.

More recently J. Hájek [5] and the present author [7] have introduced another type of regularity assumption which occurs naturally from certain statistical desiderata. These conditions refer to the differentiability in quadratic mean of the map $t \rightsquigarrow (dp_t/dp_\theta)^{\frac{1}{2}}$. It is rather obvious that the "differentiability in quadratic mean" condition cannot imply conditions of the Cramér type since only one differentiation operation is involved instead of two or three. Whether on the contrary Cramér's or Wald's conditions imply the "differentiability in quadratic mean" is not transparent. We shall show that this is indeed the case. In the process we shall generalize some of Cramér's results, reset a lemma of Hájek in a framework which is not restricted to translation parameters and extend in a similar manner a theorem of L. Shepp [10].

Although certain of the arguments given here remain valid for multidimensional parameter sets Θ we have limited the last section to the case of real-valued parameters. The reason is that some of the arguments about continuity of sample paths of stochastic processes do not extend directly to more than one dimension. In fact, an example of Kemperman (private communication) shows that the one-dimensional results do not remain valid without further restrictions when θ becomes three-dimensional. One could still proceed to an extension. However, the conditions under which we know at present that extension is possible do not have acceptable statistical relevance. Also they appear to be remote from the conditions which justify their existence by being necessary.

2. The Hellinger transform and statistical tests. Let P and Q be two probability measures on a σ-field α. Let $\mu = P+Q$ and let f and g be the densities $f = dP/d\mu$ and $g = dQ/d\mu$. The L_1-norm of the difference $P-Q$ is $\|P-Q\| = \int|f-g|\,d\mu = 2[1-\|P\wedge Q\|]$, where $P\wedge Q$ is the minimum of the two measures P and Q; that is, the measure which has density $f\wedge g$ with respect to μ. The Hellinger transform of the pair $\{P, Q\}$ is the function φ defined on [0, 1] by $\varphi(\alpha) = \|(dP)^{1-\alpha}(dQ)^{\alpha}\| = \int f^{1-\alpha}g^{\alpha}\,d\mu$. The value $\rho(P,Q) = \varphi(\frac{1}{2}) = \int(dP\,dQ)^{\frac{1}{2}}$ is also called the affinity between P and Q. The Hellinger distance $H(P,Q)$ is defined by $H^2(P,Q) = \int((dP)^{\frac{1}{2}}-(dQ)^{\frac{1}{2}})^2 = \int(f^{\frac{1}{2}}-g^{\frac{1}{2}})^2\,d\mu = 2[1-\rho(P,Q)]$. It follows that

$$H^2(P,Q) \leqq \|P-Q\| \leqq H(P,Q)(4-H^2)^{\frac{1}{2}} \leqq 2H(P,Q).$$

Thus P and Q are disjoint if and only if $\|P-Q\| = 2$ or equivalently $H^2(P,Q) = 2$. The metric H and the L_1-norm define the same uniformity on the set of probability measures.

Consider the problem of testing P against Q. There exists a test which minimizes the sum of the probabilities of error. This minimum sum of errors is easily seen to be $\|P\wedge Q\|$. Thus the L_1-norm has a direct statistical interpretation. It measures how easily one can test between P and Q. The Hellinger function φ has the following interpretation. Let Λ be the logarithm of likelihood ratio $\Lambda = \log g/f = \log dQ/dP$, with usual conventions for infinite values. Then for every $\alpha\in(0, 1)$ the value $\varphi(\alpha) = \|(dP)^{1-\alpha}(dQ)^{\alpha}\|$ is simply the value of the Laplace transform $Ee^{\alpha\Lambda}$ for an expectation taken under P. Hölder's inequality implies that $\log\varphi(\alpha)$ is convex in α. It is easily seen that $\lim_{\alpha\to 1,\,\alpha<1}\varphi(\alpha)$ is the mass of the part of Q which is dominated by P. Similarly $\lim_{\alpha\to 0,\,\alpha>0}\varphi(\alpha)$ is the mass of the part of P which is dominated by Q.

Suppose now that P is the direct product $P = X_j p_j$ of a finite or infinite family of probability measures p_j with p_j carried by a σ-field α_j on a set $\mathfrak{X}_j$. Suppose similarly that $Q = X_j q_j$ with q_j carried by the same α_j as p_j. For each pair (p_j, q_j) we have the Hellinger transform $\varphi_j(\alpha) = \|(dp_j)^{1-\alpha}(dq_j)^{\alpha}\|$. Let φ be the Hellinger transform relative to (P, Q). Then $\varphi(\alpha) = \prod_j \varphi_j(\alpha)$. One concludes easily, using the convexity of $\log \varphi_j$, that either $\varphi(\frac{1}{2}) = \prod_j \varphi_j(\frac{1}{2}) = 0$ and therefore the measures P and Q are disjoint, or $\varphi(\frac{1}{2}) > 0$, and then the mass of P which is Q dominated is precisely the product of $\prod_j m_j$ where m_j is the mass of p_j which is q_j dominated. To prove it, note that $\psi_n(\alpha) = \prod_{j\leq n}\varphi_j(\alpha)$ is a decreasing sequence of functions and that for $\alpha \in [0, \frac{1}{2})$

$$\log \varphi_j(\alpha) \geqq 2(1-\alpha)\log \varphi_j(\tfrac{1}{2});$$

therefore $\sum_{m\leq j<n}|\log \varphi_j(\alpha)| \leqq 2(1-\alpha)\sum_{m\leq j<n}|\log \varphi_j(\frac{1}{2})|$ for every $\alpha \in [0, \frac{1}{2})$.

The alternative "either $\varphi(\frac{1}{2}) = \rho(P, Q) = 0$ or the dominated masses of product are the product of dominated masses on components" will be referred to as Kakutani's alternative, from the theorem by this author [6].

Although the *statistical* interpretation of L_1-norm in terms of probabilities of errors renders the use of L_1-norm very logical, the relations between L_1-norm of components and L_1-norms of products are neither simple nor very informative. Thus one is led to use instead the Hellinger transform, or when this is satisfactory, the affinity or the Hellinger distance. In particular, the relation $\varphi(\frac{1}{2}) = \prod_j \varphi_j(\frac{1}{2})$ gives

$$H^2(P, Q) = 2\{1 - \textstyle\prod_j [1 - \frac{1}{2}H^2(p_j, q_j)]\}.$$

In all cases where the components (p_j, q_j) are such that $H^2(p_j, q_j)$ is "small" this leads to the consideration of $\sum_j H^2(p_j, q_j)$ instead of $\prod_j[1-\frac{1}{2}H^2(p_j\, q_j)]$ according to the usual inequalities. Finally this leads to the introduction of stochastic processes as follows.

Let Θ be an arbitrary set. For each $\theta \in \Theta$ let p_θ be a probability measure on a σ-field α. Let $C(s, t)$ be the affinity $C(s, t) = \int (dp_s\, dp_t)^{\frac{1}{2}}$. This function is a covariance kernel on the set Θ. Therefore, one can define real stochastic processes $t \rightsquigarrow \xi(t)$ such that $E\xi(s)\xi(t) = C(s, t)$ or equivalently $E|\xi(s) - \xi(t)|^2 = H^2(p_s, p_t)$. According to the preceding discussion, the study of product $P_{\theta,n} = X_j p_{\theta,n,j}$ can thus be reduced to a large extent to the study of series of independent processes $\xi(\theta, n, j)$ with

$$E\,|\xi(s,n,j) - \xi(t,n,j)|^2 = H^2(p_{s,n,j}, p_{t,n,j}).$$

At least one can effect such a reduction for the study of separation of measures. More precise relations between the above processes or some closely related ones will be given in the next section.

3. Condition (*l*) and some of its implications. In this section it will be assumed that Θ is a measurable subset of the k-dimensional Euclidean space R_k. For each $\theta \in \Theta$ we shall assume a given probability measure p_θ in some space $\{\mathfrak{X}, \alpha\}$. We shall denote $h(s, t)$ the Hellinger distance $h(s, t) = H(p_s, p_t)$ and suppose that $t \rightsquigarrow \xi(t)$ is a second order process such that

$$E\,|\xi(s)-\xi(t)|^2 = h^2(s,t).$$

DEFINITION. The process ξ satisfies condition (l) at the point $\theta \in \Theta$ if

$$\limsup_{|\tau|\to 0} |\tau|^{-1}\,||\xi(\theta+\tau)-\xi(\theta)|| < \infty$$

for the quadratic norm $||\xi(s)-\xi(t)||^2 = h^2(s, t)$ and for arguments $\theta+\tau \in \Theta$.

As was indicated in the preceding section, condition (l) at θ is equivalent to the statement that the products $\{P_{\theta+\tau_n/n^{\frac{1}{2}},\, n}\}$ and $\{P_{\theta,\, n}\}$ do not separate entirely as long as τ_n keep bounded.

A condition related to (l) but definitely stronger is the local Lipschitz condition (L) that there is an $\varepsilon > 0$ and a $C < \infty$ such that if $|s-\theta|+|t-\theta| < \varepsilon$ then

$$||\xi(s)-\xi(t)|| \leqq C||s-t||.$$

However (l) nearly implies (L) in the following sense.

LEMMA 1. *Assume that Θ is the intersection of a closed set with an open set of R_k and that (l) is satisfied at all points of Θ then the set (B) of points at which (L) is not satisfied is a closed set without interior points in the relative topology of Θ.*

PROOF. Condition (l) implies the continuity of the map $t \rightsquigarrow \xi(t)$ to the Hilbert space of square integrable variables, hence the continuity of the map $(s,t) \rightsquigarrow h(s, t)$. Let A_n be the set of points $\theta \in \Theta$ defined by $A_n = \{\theta \in \Theta;\ \sup_t [h(t, \theta)||t-\theta||^{-1};\ |t-\theta| \leqq n^{-1}] \leqq n\}$. By construction A_n is closed (in Θ). Let $A_n{}^0$ be the interior of A_n. A point θ satisfies (L) if and only if $\theta \in \bigcup_n A_n{}^0$. However $\bigcup A_n = \Theta$. Thus $B \subset \bigcup (A_n \backslash A_n{}^0)$ is of the first category. Furthermore, let U be any open subset of Θ (in the relative topology). If h restricted to U satisfies either (l) or (L) at some $\theta \in U$ then h satisfies the same condition at θ relatively to the whole of Θ. Thus $B \cap U$ is of the first category in U. The result follows since U is not of the first category in itself.

A deeper and more important result is the following.

THEOREM 1. *Let S be a measurable set of points $\theta \in \Theta$ at which the process ξ satisfies condition (l). Then the process ξ is differentiable in quadratic mean at all points of S except perhaps those of a subset which has Lebesgue measure zero.*

The proof will only be sketched since the result is essentially well known. A basic remark is the following lemma which is used in Saks [9] but is not isolated there.

LEMMA 2. *Let ξ be a Banach valued function defined on the measurable set Θ. Let A be a measurable subset of Θ such that for every $x \in A$ there is an $\varepsilon_x > 0$ and a $K < \infty$ such that if $y \in A$, $z \in \Theta$ and $||y-x|| < \varepsilon_x$ and $||z-x|| < \varepsilon_x$ then*

$$||\xi(y)-\xi(z)|| \leqq K\,||y-z||.$$

If ξ possesses an approximate derivative at almost all points of A then ξ possesses a derivative at almost all points of A.

This lemma can be applied to the present case by noting that the set

$$A_n = \{\theta \in S : \sup_{t \in \Theta}[h(t,\theta)|t-\theta|^{-1}; |t-\theta| < n^{-1}] \leqq n\}$$

is a measurable set. Furthermore, if $\theta \in A_n$, $s \in A_n$, $t \in \Theta$ and if $2|\theta - s| < n^{-1}$ and $2|t-\theta| < n^{-1}$ then $|s-t| < n^{-1}$ and therefore $h(t, s) \leqq n|t-s|$. Thus we are reduced to proving existence of *approximate* derivatives on sets such as A_n. Retaining only those points of A_n which are points of density unity of A_n eliminates only a set of Lebesgue measure zero. Thus it is sufficient to prove that *restricted* to A_n the function ξ has approximate derivates almost everywhere. One can also cover A_n by a countable family of sets having diameter smaller than n^{-1}. Thus it is sufficient to show that if ξ is defined on a measurable set and satisfies a Lipschitz condition on that set it is almost everywhere approximately differentiable on that set.

This is proved by induction on the number of dimensions in Saks [9] page 300. It remains therefore only to prove the starting point of the induction assumption namely that if S is a measurable bounded subset of the line and if $s \rightleftarrows \xi(s)$ is defined on S and Lipschitzian there then ξ is almost everywhere differentiable. Now such a ξ can be extended to be a Lipschitz function on an *interval* containing S. The result is then a consequence of a theorem of Gel'fand [4] according to which if a function $s \rightleftarrows \xi(s)$ to a Banach space possesses almost everywhere an almost separably valued bounded *weak* derivative it admits the same function for *strong* derivative almost everywhere.

Returning to the measures $\{p_\theta; \theta \in \Theta\}$ let us consider now other processes related to the process $\xi(t)$. For this we shall make the domination assumption that there is a probability measure μ which is such that $\mu(A) = 0$ is equivalent to $\sup_\theta p_\theta(A) = 0$.

We note in passing the following easy result.

LEMMA 2. *If the process ξ is continuous in quadratic mean and therefore in particular if ξ satisfies (l) at all $\theta \in \Theta$ then there is a probability measure μ such that $\mu(A) = 0$ if and only if $p_\theta(A) = 0$ for all $\theta \in \Theta$.*

Indeed continuity of $\theta \rightleftarrows \xi(\theta)$ implies continuity of the map $\theta \rightleftarrows P_\theta$ for the L_1-norm $\|p_\theta - p_t\|$. Hence $\{p_\theta : \theta \in \Theta\}$ possesses a countable dense subset $\{p_{\theta_j}\}$. It is sufficient to take $\mu = \sum 2^{-j} p_{\theta_j}$.

From this point on we shall assume *that ξ is the process obtained as follows*: Take the Radon–Nikodym derivative $dp_\theta/d\mu$ and its square root $(dp_\theta/d\mu)^{\frac{1}{2}}$. This is a measurable function in which one can substitute a variable ω whose distribution is given by the dominating measure μ. It is quite clear that if $\theta \rightleftarrows \xi(\theta)$ is the process so obtained then $E|\xi(\theta)|^2 = 1$ and $E|\xi(\theta) - \xi(t)|^2 = h^2(\theta, t)$.

For a given $\theta \in \Theta$ let us define other processes $X(t)$ and $Y(t)$ as follows. The process $X(t)$ is given by $X(t) = (dp_t/dp_\theta)^{\frac{1}{2}} - 1$ where one substitutes in the function dp_t/dp_θ a variable x whose distribution is given by p_θ. If necessary we shall indicate this dependence on θ by using the symbol $X_\theta(t)$ instead of $X(t)$. The process $Y(t)$, or more specifically $Y_\theta(t)$, is defined by $Y_\theta(t) = X_\theta(t) - E_\theta X_\theta(t)$.

Let J_θ be the indicator of the set of values of x where $\xi(\theta) \neq 0$. The affinity $\rho(p_s, p_t)$ may be written

$$\rho(p_s, p_t) = \int (dp_s\, dp_t)^{\frac{1}{2}} = \int J_\theta (dp_s\, dp_t)^{\frac{1}{2}} + \int (1-J_\theta)(dp_s\, dp_t)^{\frac{1}{2}}.$$

The first integral can now be rewritten

$$\int J_\theta (dp_s\, dp_t)^{\frac{1}{2}} = E_\theta[1+X_\theta(t)][1+X_\theta(s)].$$

For the second integral we shall use the notation

$$\int (1-J_\theta)(dp_s\, dp_t)^{\frac{1}{2}} = \beta_\theta(s,t).$$

It follows that

$$EX(t) = \rho(p_t, p_\theta) - 1 = -\tfrac{1}{2}h^2(\theta, t)$$

$$EX^2(t) = 2[1-\rho(p_t, p_\theta)] - \beta_\theta(t,t)$$

$$\operatorname{Var} X(t) = [1-\rho^2(p_t, p_\theta)] - \beta_\theta(t,t).$$

Also

$$\begin{aligned}\operatorname{Cov}[X(s), X(t)] &= E[Y(s)Y(t)] \\ &= \rho(p_s, p_t) - \rho(p_t, p_\theta)\rho(p_s, p_\theta) - \beta_\theta(s,t).\end{aligned}$$

We shall be interested below in cases where the process $X_\theta(t)$ is differentiable in quadratic mean at $t = \theta$. Suppose then that it is so and that the derivative in quadratic mean is $X'(\theta)$. Then the following relations hold:

(1)
$$\begin{aligned}EtX'(\theta) &= -\tfrac{1}{2}\lim_{\varepsilon\to 0}\varepsilon^{-1}h^2(\theta, \theta+\varepsilon t) \\ &= -\tfrac{1}{2}\lim_{\varepsilon\to 0}\varepsilon^{-1}\beta_\theta[\theta+\varepsilon t, \theta+\varepsilon t].\end{aligned}$$

(2)
$$\begin{aligned}EtX'(\theta)sX'(\theta) &= \lim_{\varepsilon\to 0}\varepsilon^{-2}EX(\theta+\varepsilon s)X(\theta+\varepsilon t) \\ &= \lim_{\varepsilon\to 0}\varepsilon^{-2}\{[\rho[\theta+\varepsilon s, \theta+\varepsilon t] - \rho(\theta+\varepsilon s, \theta) - \rho(\theta, \theta+\varepsilon t) + 1] \\ &\qquad - \beta_\theta(\theta+\varepsilon s, \theta+\varepsilon t)\}. \\ &= \lim_{\varepsilon\to 0} 2\varepsilon^{-2}\{h^2(\theta, \theta+\varepsilon s) + h^2(\theta, \theta+\varepsilon t) - h^2(\theta+\varepsilon s, \theta+\varepsilon t) \\ &\qquad - 2\beta(\theta+\varepsilon s, \theta+\varepsilon t)\}.\end{aligned}$$

From these relations one can easily deduce that if $X'(\theta)$ exists and if ξ satisfies condition (l) at θ then $EX'(\theta) = 0$. More specifically one can obtain the following result.

LEMMA 3. *Assume that the process $t \rightsquigarrow \xi(t)$ is differentiable in quadratic mean at $t = \theta$. Let $\xi'(\theta)$ be the derivative and let $\Gamma(\theta)$ be the matrix $\Gamma(\theta) = E\xi'(\theta)[\xi'(\theta)]^T$. Then both $X_\theta(t)$ and $Y_\theta(t)$ are differentiable in quadratic mean at $t = \theta$. Also $EX_\theta'(\theta) = 0$ and $X_\theta'(\theta) = Y_\theta'(\theta)$. Finally, if $\gamma(\theta)$ is the matrix $EX'(\theta)[X'(\theta)]^T$ then*

$$s[\Gamma(\theta) - \gamma(\theta)]t^T = \lim_{\varepsilon\to 0}\varepsilon^{-2}\beta_\theta[\theta+\varepsilon s, \theta+\varepsilon t].$$

The difference $\Gamma(\theta) - \gamma(\theta)$ is a positive semidefinite matrix.

PROOF. We have already remarked that differentiability of X and condition (l) imply that $EX'(\theta) = 0$. Since $Y(t) = X(t)+\frac{1}{2}h^2(\theta, t)$ the relation between Y' and X' will also follow if we prove only that the differentiability of ξ implies that of X.

To show this let J_θ be the indicator of the set where $\xi(\theta) \neq 0$. One can then write

$$\xi(t)-\xi(\theta) = J_\theta[\xi(t)-\xi(\theta)]+(1-J_\theta)\xi(t)$$

and

$$[\xi(t)-\xi(\theta)-(t-\theta)\xi'(\theta)] = J_\theta[\xi(t)-\xi(\theta)-(t-\theta)\xi'(\theta)]+(1-J_\theta)[\xi(t)-(t-\theta)\xi'(\theta)].$$

Divide by $|t-\theta|$ and take expectations of the squares. Since the two terms on the right are disjoint, both norms

$$|t-\theta|^{-1}\,||J_\theta[\xi(t)-\xi(\theta)-(t-\theta)\xi'(\theta)]||$$

and

$$|t-\theta|^{-1}\,||(1-J_\theta)[\xi(t)-(t-\theta)\xi'(\theta)]||$$

must converge to zero. The first term is the norm obtained from the integral

$$\begin{aligned} &|t-\theta|^{-2}\int J_\theta\,|\xi(t)-\xi(\theta)-(t-\theta)\xi'(\theta)|^2\,d\mu \\ &= |t-\theta|^{-2}\int J_\theta\left|\frac{\xi(t)}{\xi(\theta)}-1-(t-\theta)\frac{\xi'(\theta)}{\xi(\theta)}\right|^2\xi^2(\theta)\,d\mu \\ &= |t-\theta|^{-2}E_\theta\,|X(t)-(t-\theta)X'(\theta)|^2 \end{aligned}$$

for $X'(\theta) = \xi'(\theta)/\xi(\theta)$. In other words, as functions on the probability space, the derivatives ξ' and X' satisfy the relation:

$$\xi'(\theta) = \xi(\theta)X'(\theta)+(1-J_\theta)\xi'(\theta).$$

Since $(1-J_\theta)\xi(\theta) = 0$ the desired relations follow immediately.

Note that since $\Gamma(\theta)-\gamma(\theta)$ is semidefinite, to show that $\Gamma(\theta)-\gamma(\theta) = 0$ at a particular point θ it will be (necessary and) sufficient to show that for each fixed t one has

$$\lim_{\varepsilon\to 0}\varepsilon^{-2}\beta_\theta(\theta+\varepsilon t, \theta+\varepsilon t) = 0.$$

We shall now show that under condition (l) this is almost everywhere true.

PROPOSITION 1. *Assume that the process $t \rightsquigarrow \xi(t)$ is differentiable in quadratic mean at almost all points of the measurable set Θ. Then*

$$\lim_{|\tau|\to 0}|\tau|^{-2}\beta_\theta[\theta+\tau, \theta+\tau] = 0$$

for all $\theta\in\Theta$ except perhaps those of a subset which has Lebesgue measure zero.

PROOF. For simplicity let us write $\varphi(\theta, t) = \beta_\theta(t, t)$. Thus $\varphi(\theta, t)$ is simply the mass of the part of p_t which is p_θ singular. We have seen above that when the process ξ is differentiable at θ then $\lim|\tau|^{-2}\varphi[\theta, \ \theta+\tau] \leqq \sup\{u\Gamma(\theta)u^T, \ ||u|| \leqq 1\} < \infty$. In particular, $\varphi(\theta, \theta+\tau)\to 0$. Also, since $EX'(\theta) = 0$, $\lim\varepsilon^{-1}h^2(\theta, \theta+\varepsilon\tau) = 0$ so that $\lim \varepsilon^{-1}\varphi(\theta+\varepsilon\tau, \theta) = 0$ and more generally $\lim_{|\tau|\to 0} |\tau|^{-1}\varphi[\theta+\tau, \theta] = 0$. Let us show in fact that at such a point θ one has always

$$\lim_{|\tau|\to 0} |\tau|^{-2}\varphi[\theta+\tau, \theta] = 0.$$

For this purpose write

$$r(\theta, t) = |t-\theta|^{-1} \|\xi(t)-\xi(\theta)-(t-\theta)\xi'(\theta)\|.$$

If J_t is as before the indicator of the set where $\xi(t) \neq 0$, the part of p_θ which is p_t singular contributes to $r^2(\theta, t)$ a term of the type

$$|t-\theta|^{-2} EJ_t |\xi(\theta)+(t-\theta)\xi'(\theta)|^2.$$

Thus

$$r(\theta, t) \geqq |t-\theta|^{-1} |EJ_t|\xi(\theta)|^2|^{\frac{1}{2}} - |EJ_t|\xi'(\theta)|^2|^{\frac{1}{2}}.$$

Since as $t \to \theta$ $EJ_t \to 0$ the last term on the right tends to zero. Therefore the same is true of the first term which proves our assertion.

To go further, note that ξ is continuous at almost every point of Θ. Let Θ_0 be the set of points of continuity and let

$$r_n(\theta) = \sup_t \{r(\theta, t); t\in\Theta_0, |t-\theta| \leqq n^{-1}\}.$$

This is a measurable function of $\theta\in\Theta_0$. By assumption $r_n(\theta) \to 0$ for almost every $\theta\in\Theta_0$; hence by Egorov's theorem, r_n converges to zero uniformly on certain compacts of Θ_0 which cover all of Θ_0 except a Lebesgue null set. Let us consider then a compact $K\subset\Theta_0$ such that $r_n = \sup[r_n(\theta), \theta\in K] \to 0$ and such that $\xi'(\theta)$ be continuous when restricted to K. If θ_n and t_n are both elements of K and $|\theta_n - t_n| < n^{-1}$ we can write again

$$r_n \geqq r(\theta_n, t_n) \geqq |t_n-\theta_n|^{-1} |EJ_{t_n|}\xi(\theta_n)|^2|^{\frac{1}{2}} - E|J_{t_n}|\xi'(\theta_n)|^2|^{\frac{1}{2}}.$$

We can assume that $\theta_n \to \theta_0\in K$. In this case $\xi'(\theta_n) \to \xi'(\theta_0)$ in quadratic mean. Also $EJ_{t_n} \to 0$ since $EJ_{t_n} \leqq h^2(t_n, \theta_n)$. Thus the last term still tends to zero. The same is therefore true of the first term.

To summarize, if K is a compact such that $\sup[r_n(\theta); \theta\in K] \to 0$ and such that ξ' restricted to K is continuous, then

$$\sup_{\theta,t}\{\varphi(\theta, t)|t-\theta|^{-2}; \theta\in K, t\in K, (\theta-t) < n^{-1}\} \to 0.$$

This proves the desired result, since it implies $\gamma(\theta) = \Gamma(\theta)$ at least for all points of density unity of K.

LEMMA 4. *Assume that the process ξ is differentiable in quadratic mean at θ. Let $\{X_j, j = 1, 2, \cdots\}$ be independent copies of the process $t \rightleftarrows X(t)$ induced by the measure p_θ. Let X_j' be the derivative of X_j at $t = \theta$ and let V_n be the random variable*

$$V_n = n^{-\frac{1}{2}}\sum_{j=1}^n X_j'(\theta).$$

If τ_n is a bounded sequence of vectors then

$$\sum_{j=1}^{n} X_j(\theta+\tau_n/n^{\frac{1}{2}})-\tau_n V_n+\tfrac{1}{2}\tau_n\Gamma(\theta)\tau_n^T$$

converges to zero in quadratic mean as $n\to\infty$.

PROOF. Consider first the processes $Y_j(t)=X_j(t)-EX_j(t)$. According to the preceding lemma, Y_j is differentiable in quadratic mean at $t=\theta$ and the derivative is $X_j'(\theta)$. It follows that

$$Y_j(\theta+t)=tX_j'(\theta)+|t|\,R_j(t)$$

where $ER_j=0$ and where $ER_j^2(t)\to 0$ if $t\to 0$. This gives

$$\sum_{j\leqq n} Y_j(\theta+\tau/n^{\frac{1}{2}})=\tau V_n+|\tau|\,n^{-\frac{1}{2}}\sum_{j\leqq n} R_j(\tau/n^{\frac{1}{2}}).$$

The last sum on the right has expectation zero and a variance equal to the sum of the variances of its components.

$$E\,|n^{-\frac{1}{2}}\sum_{j\leqq n} R_j(\tau/n^{\frac{1}{2}})|^2=ER_1^2(\tau/n^{\frac{1}{2}}).$$

This tends to zero as $n\to\infty$ as long as τ stays bounded. To complete the proof it is sufficient to evaluate the sum $\sum_j EX_j(\theta+\tau/n^{-\frac{1}{2}})$. We have seen that this is equal to $-n\frac{1}{2}h^2(\theta,\theta+\tau/n^{\frac{1}{2}})$ or equivalently to

$$-(n/2)E\,|\xi(\theta+\tau/n^{\frac{1}{2}})-\xi(\theta)|^2.$$

Since ξ is differentiable at θ this tends to $-\frac{1}{2}E|\tau\xi'(\theta)|^2$ provided only that τ remains bounded.

REMARK. The above argument remains valid if one assumes only that X is differentiable in quadratic mean at $t=\theta$, provided that, for instance,

$$nh^2(\theta,\theta+\tau_n/n^{\frac{1}{2}})$$

remains bounded. However, in this case the statement of the lemma must be modified and replaced by the assertion that

$$\sum_{j=1}^{n} X_j(\theta+\tau_n/n^{\frac{1}{2}})-\tau_n V_n+\tfrac{1}{2}h^2(\theta,\theta+\tau_n/n^{\frac{1}{2}})$$

tends to zero in quadratic mean.

COROLLARY. *Under the conditions of Lemma 4 the process $\tau \rightleftarrows \sum_{j=1}^{n} X_j(\theta+\tau/n^{\frac{1}{2}})$ has finite dimensional distributions which converge to that of a normal process $\tau\rightleftarrows Z(\tau)$ with expectation $EZ(\tau)=-\frac{1}{2}\tau\Gamma(\theta)\tau$ and covariance kernel*

$$E[Z(s)-EZ(s)]Z(t)=s\gamma(\theta)t^T.$$

In particular if $\tau_n\to\tau$ the sequences $\{P_{\theta,n}\}$ and $\{P_{\theta+\tau_n/n^{1/2}}\}$ are contiguous if and only if

$$\tau\gamma(\theta)\tau^T=\tau\Gamma(\theta)\tau^T.$$

This is quite obvious.

Instead of the sum $\sum_{j \leqq n} X_j(\theta + \tau/n^{\frac{1}{2}})$ one can also consider the logarithm of likelihood ratio obtained by taking independent identically distributed observations.

Letting $\Lambda_n(\theta, t) = \log dP_{t,n}/dP_{\theta,n}$ be the logarithm of likelihood ratio, one can take Λ_n equal to

$$\Lambda_n(\theta, t) = 2\sum_{j=1}^n \log[1 + X_j(t)]$$

for the distributions induced by p_θ. The limiting behavior of $\Lambda_n(\theta, t)$ is related to that of $\sum_j X_j(t)$ according to the prescription given by the following result.

LEMMA 5. *Assume that the process $t \rightleftarrows X(t)$ is differentiable in quadratic mean at $t = \theta$ and that ξ satisfies condition (l) at θ. Then, for every bounded sequence τ_n the difference*

$$\tfrac{1}{2}\Lambda_n(\theta, \theta + \tau_n/n^{\frac{1}{2}}) - \sum_{j=1}^n X_j(\theta + \tau_n/n^{\frac{1}{2}}) + \tfrac{1}{2}\tau_n \gamma(\theta)\tau_n$$

tends to zero in $P_{\theta,n}$ probability as $n \to \infty$.

This follows from the asymptotic normality of $\sum X_j(\theta + \tau_n/n^{\frac{1}{2}})$ and the usual Taylor expansion argument.

In the case where ξ is differentiable in quadratic mean this lemma shows that the limiting distribution of $\Lambda_n(\theta, \theta + \tau/n^{\frac{1}{2}})$ is a Gaussian distribution with variance $4\tau\gamma(\theta)\tau^T$ and with expectation equal to $-\tau[\Gamma(\theta) + \gamma(\theta)]\tau^T$.

Another aspect of the implications of condition (l) is related to results of L. Shepp in [10]. Theorems 1 and 3 of [10] are implied by the following proposition.

PROPOSITION 2. *Assume that θ is such that for any $t \in \Theta$ the affinity $\rho(p_\theta, p_t)$ is not zero. Consider the measure P direct product of a countable set of copies of p_θ. For a sequence $T = \{t_n\}$ of vectors such that $\theta + t_n \in \Theta$ for all n let Q_T be the direct product of the measures $p_{\theta + t_n}$.*

If the process ξ satisfies condition (l) at θ then for any sequence $T = \{t_n\}$ such that $\sum |t_n|^2 < \infty$ the measures P and Q_T are nondisjoint and the masses of their dominated parts are the product of the masses of the dominated parts of the components.

Conversely, if $\rho(P, Q_T) > 0$ for every sequence $T = \{t_n\}$ such that $\sum |t_n|^2 < \infty$ then condition (l) is satisfied at θ.

PROOF. Let $\rho_n = \rho(p_\theta, p_{\theta + t_n})$ and let $h_n^2 = 2(1 - \rho_n) = H^2(p_\theta, p_{\theta + t_n})$. Since we have assumed that $\rho_n > 0$ the product $\prod \rho_n$ is convergent in $(0, \infty)$, that is $\rho(P, Q_T) > 0$ if and only if $\sum_n h_n^2 < \infty$. Thus, to prove the first statement it is sufficient to apply Kakutani's alternative theorem (Section 2) and note that if

$$b(\theta) = \limsup_{|\tau| \to 0} |\tau|^{-1} \|\xi(\theta + \tau) - \xi(\theta)\|$$

then

$$\sum_{n \geqq m} h_n^2 \leqq [b^2(\theta) + 1] \sum_{n \geqq m} |t_n|^2$$

for m sufficiently large.

Conversely, suppose that $\sum |t_n|^2 < \infty$ implies $\sum h_n^2 < \infty$. Let $F(u)$ be the number

$$F(u) = \sup_t \{h^2(\theta, \theta + t);\ |t|^2 = u\}.$$

Suppose that there is a sequence $u_n > 0$ such that $u_n \to 0$ but $\lim_n F(u_n)/u_n = \infty$. Extracting subsequences, if necessary, one may assume that the sequence u_n is such that

(a) the sequence u_n decreases and $10\, u_{n+1} \leqq u_n$.
(b) $F(u_{n+1})/u_{n+1} > 5F(u_n)/u_n$.

Construct a sequence of numbers r_n as follows. Take $r_1 = 1$ and let $r_{n+1} \geqq 2$ be such that

(c) $r_{n+1}\,(u_{n+1}/u_n) \leqq \frac{4}{5}$.
(d) $r_{n+1}\,[F(u_{n+1})/F(u_n)] \geqq \frac{6}{5}$.

These inequalities can be written

$$\tfrac{6}{5}F(u_n)/F(u_{n+1}) \leqq r_{n+1} \leqq \tfrac{4}{5}u_n/u_{n+1}.$$

Thus there is always some number $r_{n+1} \geqq 8$ which satisfies the relation. Let k_n be the integer part of $\prod_{j=1}^{n} r_j$. Let v_n be a vector such that $|v_n|^2 = u_n$ and

$$h^2(\theta, \theta + v_n) \geqq F(u_n) - 2^{-n} k_n^{-1}.$$

Form a sequence t_n by taking $t_1 = v_1$, and $t_2, t_3, \cdots, t_{1+k_2}$ equal to v_2 and so forth, so that the value $t = v_n$ appears k_n times. Then $\sum |t_j|^2 = \sum_n k_n u_n$ still converges. However

$$\textstyle\sum_j h^2(\theta, \theta + t_j) \geqq -1 + \sum k_n F(u_n)$$

is infinite. Thus if condition (l) is not satisfied there are sequences $T = \{t_n\}$ with $\sum |t_n|^2 < \infty$ but $\rho(P, Q_T) = 0$.

In the case considered by Shepp the measures p_t are obtained by shifting the measure p_θ by the amount $t - \theta$. In such a case the kernel $C(s, t) = \rho(p_s, p_t)$ is the covariance of a stationary process. Thus, if condition (l) holds at one value of θ it holds for every θ. The process is then differentiable in quadratic mean almost everywhere; hence everywhere because of the shift invariance. Thus in the shift situation one can replace condition (l) by differentiability in quadratic mean. This is essentially the statement of Lemma 3 of [10].

4. Some other differentiability conditions. In previous papers [7], [8], this author has studied certain statistical conditions which were called there asymptotic differentiability conditions. Several equivalent forms are given in the papers mentioned above. One possible system of assumptions is the following, where we denote again the logarithm of likelihood ratio $\log(dP_{t,n})/(dP_{\theta,n})$ by $\Lambda_n(\theta, t)$, and where it is still assumed that Θ is a subset of the k-dimensional Euclidean space R_k.

(D1) If $\{|t_n|\}$ is a bounded sequence then the sequences $\{P_{\theta,n}\}$ and $\{P_{\theta + t_n/n^{\frac{1}{2}}, n}\}$ are contiguous.

(D2) There exist random vectors V_n and a function $t \rightarrow A(t)$ such that if $|t_n|$ is bounded

$$\Lambda_n(\theta, \theta + t_n/n^{\frac{1}{2}}) - t_n V_n + A(t_n)$$

tends to zero in $P_{\theta,n}$ probability.

(D3) The sequence of distributions $\mathscr{L}(V_n \mid \theta)$ has a limit.

(D4) If $\mathscr{L}\{\Lambda_n(\theta, \theta+t_n/n^{\frac{1}{2}} \mid \theta\}$ possesses a limit and if $\sup|t_n| < \infty$ the limiting distribution is Gaussian.

Let us note in passing that (D3) is almost redundant, since it is implied by the combination (D1)(D2) unless the closed cone tangent to Θ at θ does not have interior points.

Condition (D1) alone is already very much stronger than the assumption that our process ξ satisfies condition (l) at θ. In fact, one can show that (D1) is equivalent to the following set of conditions in which $X(t)$ is the process described in Section 3.

(D1, a) If $\sup|t_n| < \infty$ then the sequence of distributions $\mathscr{L}[\sum_{j\leqq n}X_j(\theta+t_n/n^{\frac{1}{2}}]$ is relatively compact.

(D1, b) Assuming $\sup|t_n| < \infty$ if the sequence $F_n = \mathscr{L}[\sum_{j\leqq n}X_j(\theta+t_n/n^{\frac{1}{2}})]$ converges to F then the variance of the limiting distribution F is the limit of the variances of the F_n.

(D1, c) Let $\varphi(s, t)$ be the mass of the part of p_t which is p_s singular. Then, if $\sup|t_n| < \infty$

$$\lim n[\varphi(\theta, \theta+t_n/n^{\frac{1}{2}})+\varphi(\theta+t_n/n^{\frac{1}{2}}, \theta)] = 0.$$

It has been shown in [7] but follows easily from the above decomposition of (D1) that the combination (D1)(D2)(D3) implies that the following (C1) is satisfied.

(C1) Let $C(s, t) = \rho(p_s, p_t)$ and assume $s_n \to s$ and $t_n \to t$. Then

$$\lim_{\varepsilon\to 0}\varepsilon^{-2}\{C(\theta+\varepsilon s_n, \theta+\varepsilon t_n)-C(\theta+\varepsilon s_n, \theta)-C(\theta, \theta+\varepsilon t_n)+C(\theta, \theta)\}$$

exists and is finite.

If in addition (D4) is satisfied the limit in Condition (C1) must be of the form $s\Gamma(\theta)t^T$ for a certain matrix $\Gamma(\theta)$. This will be referred to as Condition (C2).

Note that Condition (C2) looks similar to the condition of differentiability in quadratic mean at θ of the process ξ. However, the increments εs_n and εt_n involved in the expression tend to zero at the same rate so that Condition (C2) is weaker than the condition of differentiability in quadratic mean.

Since differentiability in quadratic mean of ξ at θ implies (D2), (D3), (D4), but not always (D1), one could conjecture that the combination of (D1), (D2), (D3), (D4) might imply the differentiability of ξ. We shall now give an example showing that this is not necessarily correct.

Let $\{\mathfrak{X}, \mathscr{A}, \lambda\}$ be the probability space consisting of the interval [0, 1] with its Borel sets and the Lebesgue measure. Let $\{\varphi_k\}$ be an orthonormal sequence of functions. That is $\int\varphi_k\, d\lambda = 0$ and $\int\varphi_j\varphi_h\, d\lambda = \delta_{j,k}$. Let $f(t) = |\log t|^{\frac{1}{2}}$ for $t\in[0, t_0]$, $t_0 < 1$. Construct a function $t \rightsquigarrow V(t)$ as follows. If $k \leqq f(t) < k+1$ let

$$V(t) = [(k+1)-f(t)]\varphi_k+[f(t)-k]\varphi_{k+1}.$$

Let $\gamma(t) > 0$ be the number such that

$$\gamma^2(t) = EV^2(t) = [k+1-f(t)]^2+[f(t)-k]^2 \geqq \tfrac{1}{2}.$$

Finally define a process $X(t)$ by writing

$$X(t) = t[V(t)/\gamma(t) - a(t)]$$

where $a(t) > 0$ is selected such that

$$2a(t) = t[1 + a^2(t)].$$

It is easily verifiable that such a value of $a(t)$ does exist and $t/2 < a(t) < t$. In order that $1 + X(t) \geqq 0$ it is sufficient that

$$V(t) \geqq \gamma(t)a(t) - 1/t.$$

If this positivity requirement is satisfied the process $1 + X(t)$ can be treated as the square root of a likelihood ratio for a measure p_t whose density with respect to $\lambda = p_0$ is $[1 + X(t)]^2$.

Let us first show that for this family Condition (C2) is satisfied at $t = 0$. For this purpose, take two values u and v such that $u < v$, and consider $V(u/n^{\frac{1}{2}})$ and $V(v/n^{\frac{1}{2}})$. The inequality $u < v$ implies $f(u/n^{\frac{1}{2}}) > f(v/n^{\frac{1}{2}})$. If there is a k such that $k \leqq f(v/n^{\frac{1}{2}}) < f(u/n^{\frac{1}{2}}) < k+1$, then

$$V(u/n^{\frac{1}{2}}) - V(u/n^{\frac{1}{2}}) = [f(u/n^{\frac{1}{2}}) - f(v/n^{\frac{1}{2}})][\varphi_{k+1} - \varphi_k].$$

Thus

$$E\left|V(u/n^{\frac{1}{2}}) - V(v/n^{\frac{1}{2}})\right|^2 \leqq 2\left|f(u/n^{\frac{1}{2}}) - f(v/n^{\frac{1}{2}})\right|^2.$$

Suppose, on the contrary, that $k \leqq f(v/n^{\frac{1}{2}}) < k+1 < f(u/n^{\frac{1}{2}}) < k+2$. In this case the coefficient of φ_{k+1} is equal to $1 - [(k+1) - f(v/n^{\frac{1}{2}})]$ for $V(v/n^{\frac{1}{2}})$ and to $1 - [f(u/n^{\frac{1}{2}}) - (k+1)]$ for $V(u/n^{\frac{1}{2}})$. The difference of the two is at most $f(u/n^{\frac{1}{2}}) - f(v/n^{\frac{1}{2}})$. Taking into account the contributions of φ_k and φ_{k+2} one obtains

$$E\left|V(u/n^{\frac{1}{2}}) - V(v/n^{\frac{1}{2}})\right|^2 \leqq 3\left|f(u/n^{\frac{1}{2}}) - f(v/n^{\frac{1}{2}})\right|^2.$$

For n sufficiently large one can write $0 < v < n^{\frac{1}{2}}$ and also

$$f(u/n^{\frac{1}{2}}) - f(v/n^{\frac{1}{2}}) = (\log n^{\frac{1}{2}} - \log u)^{\frac{1}{2}} - (\log n^{\frac{1}{2}} - \log v)^{\frac{1}{2}}.$$

This quantity is of the order of magnitude of $\frac{1}{2}(\log v/u)/(\log n^{\frac{1}{2}})^{\frac{1}{2}}$. Thus, if $u < v < n^{\frac{1}{2}}$ and if $\log v/u < b$, the difference $f(u/n^{\frac{1}{2}}) - f(v/n^{\frac{1}{2}})$ will be less than unity for sufficiently large n. From that point on the only possible cases are the two cases just described above, and therefore,

$$E\left|V(u/n^{\frac{1}{2}}) - V(v/n^{\frac{1}{2}})\right|^2 \leqq Cb/(\log n^{\frac{1}{2}})^{\frac{1}{2}}.$$

The covariance difference which occurs in Conditions (C1) and (C2) can be written here

$$\begin{aligned} n[C(u/n^{\frac{1}{2}}, v/n^{\frac{1}{2}}) &- C(u/n^{\frac{1}{2}}, 0) - C(0, v/n^{\frac{1}{2}}) + C(0,0)] \\ &= nE[X(u/n^{\frac{1}{2}}) - X(0)][X(v/n^{\frac{1}{2}}) - X(0)] \\ &= uvE[W(u/n^{\frac{1}{2}}) - a(u/n^{\frac{1}{2}})][W(v/n^{\frac{1}{2}}) - a(v/n^{\frac{1}{2}})] \\ &= uv\{EW(u/n^{\frac{1}{2}})W(v/n^{\frac{1}{2}}) - a(u/n^{\frac{1}{2}})a(v/n^{\frac{1}{2}})\} \end{aligned}$$

where we have written $W(t)$ for $V(t)/\gamma(t)$. The term involving the function a is smaller than uvn^{-1}. According to the above comparison of $V(u/n^{\frac{1}{2}})$ and $V(v/n^{\frac{1}{2}})$ the first expectation tends to unity. Therefore the second difference involved in (C1) tends to uv.

Let us consider now a more specific choice of the functions φ_k. For this purpose let g_m be the function defined by

$$\begin{aligned} g_m(x) &= 2^{\frac{1}{2}m} && \text{if} \quad x \in [0, 2^{-(m+1)}], \\ &= -2^{\frac{1}{2}m} && \text{if} \quad x \in [2^{-(m+1)}, 2^{-m}], \\ &= 0 && \text{otherwise.} \end{aligned}$$

Take for φ_k the function g_k. The inequality $k \leqq f(t) < k+1$ can be written $\exp[-(k+1)^2] < t \leqq \exp(-k^2)$. Thus if $k \leqq f(t) \leqq k+1$ we can write

$$t\,|V(t)| \leqq \exp(-k^2)2^{\frac{1}{2}(k+1)}.$$

From this one can easily verify that the corresponding $1+X(t)$ will be positive at least for $t < \frac{1}{2}$ and $k \geqq 2$. Writing again $W(t)$ for $V(t)/\gamma(t)$ take a fixed number $u \in (0, \infty)$. If k is such that

$$\exp[-(k+1)^2] < u/n^{\frac{1}{2}} \leqq \exp(-k^2)$$

then

$$|n^{-\frac{1}{2}}W(u/n^{\frac{1}{2}})| \leqq u^{-1}2\exp[-k^2]2^{\frac{1}{2}(k+1)}.$$

Suppose now that the W_j are independent random processes which are copies of the process W. It follows from the above inequality that $n^{-\frac{1}{2}}\sum_{j \leqq n} W_j(u/n^{\frac{1}{2}})$ has a limiting normal distribution with expectation zero and variance unity. The corresponding sum $\sum_{j \leqq n} X_j(u/n^{\frac{1}{2}})$ is equal to

$$\sum_{j \leqq n} X_j(u/n^{\frac{1}{2}}) = (u/n^{\frac{1}{2}})\sum_{j \leqq n} W_j(u/n^{\frac{1}{2}}) - un^{\frac{1}{2}}a(u/n^{\frac{1}{2}}).$$

This is also asymptotically normal with variance u^2 and with an expectation equal to the limit of $-un^{\frac{1}{2}}a(u/n^{\frac{1}{2}})$. Since a is defined by the equality $2a(t) = t[1+a^2(t)]$ one can write $a(t)/t = \frac{1}{2}[1+a^2(t)]$. Thus $\lim_{t \to 0} a(t)/t = \frac{1}{2}$. This gives

$$\lim un^{\frac{1}{2}}a(u/n^{\frac{1}{2}}) = \tfrac{1}{2}u^2.$$

In conclusion, the family of measures p_t constructed above with the particular choice of function φ_k has the following properties.

(a) Conditions (D1), (D2), (D3), (D4) are satisfied, and therefore Condition (C2) is also satisfied.

(b) The process $t \rightsquigarrow X(t)$ is not differentiable in quadratic mean at $t = 0$. In fact $t^{-1}X(t)$ tends to zero in measure, and even almost surely, but $E|t^{-1}X(t)|^2 = 1+a^2(t) \to 1$.

(c) There is an $\varepsilon > 0$ such that all the p_t, $t \in [0, \varepsilon]$ are mutually absolutely continuous.

It is true that the process $t \rightleftarrows X(t)$ has a certain number of kinks corresponding to the passage from the inequality $k \leqq f(t) < k+1$ to $k+1 \leqq f(t) < k+2$. One could avoid this by defining $V^*(t)$ according to the formula

$$V^*(t) = [1-q(f(t)-k)]\varphi_k + q[f(t)-k]\varphi_{k+1}$$

for $k \leqq f(t) < k+1$ and for a function q defined by

$$q(z) = C\int_0^z \exp\{-[x^{-1}+(1-x)^{-1}]\}\,dx$$

with $q(1) = 1$. Then proceeding just as before one obtains a process $X^*(t)$ which is now infinitely differentiable for every $t > 0$ and every x. The process is also infinitely differentiable in quadratic mean for $t > 0$. Finally $\lim t^{-m}V^*(t) = 0$ almost surely for every number m. In other words $X^*(t) = -\frac{1}{2}t^2[1+R(t)]$ for a process $t \rightleftarrows R(t)$ infinitely differentiable for $t > 0$ and such that $t^{-m}R(t) \to 0$ almost surely as $t \to 0$. In spite of this, conditions (D1) to (D4) are still satisfied with $\Lambda_n(0, t/n^{\frac{1}{2}})$ asymptotically normal with variance equal to $4t^2$.

It is clear that one could also extend the definition of the process to the interval $[-1, 0]$ so as to have a singularity at the interior of the domain of definition: just write $V[-|t|] = V[|t|]$.

The foregoing conditions (D1)–(D4) are also related to certain conditions of differentiability of the logarithm of densities dp_t/dp_θ. As a prelude to the study of the implications of conditions of the Cramér or Wald type let us consider the following system of assumptions.

Assume that the measures p_θ have densities $f(x, \theta)$ with respect to a certain σ-finite measure μ on the space $\{\mathfrak{X}, \mathscr{A}\}$. Let $\Phi(x, \theta)$ be the logarithm $\Phi(x, \theta) = \log f(x, \theta)$. The set Θ is still a subset of R_k.

ASSUMPTION A1. There is a vector valued function $\varphi(x, \theta)$, and matrix valued functions $B(x, \theta)$ and $B(x, \theta, t)$, such that

(1) $\Phi(x, \theta+t)-\Phi(x, \theta) = t\varphi(x, \theta)-\frac{1}{2}tB(x, \theta, t)t^T$.
(2) $E_\theta\varphi(x, \theta) = 0$. Also $M(\theta) = E_\theta\varphi(x, \theta)\varphi^T(x, \theta)$ exists.
(3) $\lim_{\varepsilon\to 0}\sup_{|t|<\varepsilon} E_\theta\|B(x, \theta, t)-B(x, \theta)\| = 0$.
(4) $C(\theta) = E_\theta B(x, \theta)$ exists.

We shall also need a stronger assumption as follows.

ASSUMPTION (A2). All the conditions of (A1) are satisfied but part 3 of (A1) is replaced by

$$\lim_{\varepsilon\to 0} E_\theta \sup_{|t|<\varepsilon}\|B(x, \theta, t)-B(x, \theta)\| = 0.$$

Let $\Delta_n = n^{-\frac{1}{2}}\sum_{j\leqq n}\varphi(x_j, \theta)$, the variables x_j being independent identically distributed according to the measure p_θ. Let $\Lambda_n(\theta, t)$ be the logarithm of likelihood ratio $\Lambda_n(\theta, t) = \sum_{j\leqq n}[\Phi(x_j, t)-\Phi(x_j, \theta)]$.

LEMMA 6. *Assume that condition* (A1) *is satisfied then, for any sequence* t_n *such that* $\theta + t_n/n^{\frac{1}{2}} \in \Theta$ *and* $\sup|t_n| < \infty$. *The difference*

$$\Lambda_n(\theta, \theta + (t_n/n^{\frac{1}{2}})) - t_n \Delta_n + \tfrac{1}{2} t_n C(\theta) t_n^T$$

tends to zero in probability.

PROOF. The proof is immediate. The difference in question is simply equal to

$$-n^{-1} \textstyle\sum_{j \leqq n} t_n[B(x_j, \theta, t_n/n^{\frac{1}{2}}) - C(\theta)]t_n^T = -n^{-1} \textstyle\sum_{j \leqq n} t_n[B(x_j, \theta, t_n/n^{\frac{1}{2}}) - B(x_j, \theta)]t_n^T$$
$$+ n^{-1} \textstyle\sum_{j \leqq n} t_n[B(x_j, \theta) - C(\theta)]t_n^T.$$

The last term tends to zero according to the strong law of large numbers. For the first term one can use the bound $|t_n|^2 n^{-1} \sum_{j \leqq n} \|B(x_j, \theta, t_n/n^{\frac{1}{2}}) - B(x_j, \theta)\|$. By Assumption (A1) part 3 the expectation of this tends to zero.

COROLLARY. *If* (A1) *is satisfied and* $t_n \to t$ *then* $\Lambda_n(\theta, \theta + (t_n/n^{\frac{1}{2}})$ *is asymptotically normally distributed with mean* $-\frac{1}{2}tC(\theta)t^T$ *and variance* $tM(\theta)t^T$. *The difference* $C(\theta) - M(\theta)$ *is positive semi-definite.*

The last statement arises from the fact that if Λ has the appropriate limiting distribution of Λ_n then $Ee^{\Lambda} \leqq 1$ with equality only when the sequences are contiguous.

COROLLARY. *Condition* (A1) *implies* (D2) (D3) (D4) *but not* (D1). *However,* (D1) *is implied by* (A1) *and the further requirement that* $C(\theta) = M(\theta)$.

Let us note in passing that the contiguity requirement (D1) does imply $C(\theta) = M(\theta)$ if the closed cone tangent to Θ at θ contains at least k linearly independent vectors. Also, Assumption (A1) as stated does imply that p_θ is absolutely continuous with respect to each p_t for t in a neighborhood of θ. However, it does *not* imply that p_t is absolutely continuous with respect to p_θ. The singular mass may be approximately equal to $\frac{1}{2}(t-\theta)C(\theta)(t-\theta)^T$.

It is clear that (A1) implies condition (l) at the point θ. Also (A1) implies that $\Phi(x, t)$ is differentiable in p_θ measure at θ. Therefore (A1) implies that the process $t \rightsquigarrow X_\theta(t)$ is differentiable in measure at θ. In fact, the derivative $X_\theta'(\theta)$ is easily seen to be equal to $\frac{1}{2}\varphi(x, \theta)$.

As usual, if $t_n \to t$ the asymptotic normality of $\Lambda_n(\theta, \theta + (t_n/n^{\frac{1}{2}}))$ implies the asymptotic normality of $\sum_{j \leqq n} X_j(t_n/n^{\frac{1}{2}})$ with a limiting distribution which has a variance equal to $\frac{1}{4}tM(\theta)t^T$ and an expectation equal to $-\frac{1}{4}tC(\theta)t^T + \frac{1}{8}tM(\theta)t^T$. If the process ξ is differentiable in quadratic mean at θ, the matrices C and M are related to our previous matrices γ and Γ by the relations $M = 4\gamma$ and $C = 2(\gamma + \Gamma)$. When $X(t)$ is differentiable *in measure* the following remark may be of interest.

LEMMA 7. *Suppose that there is a random vector* V *such that*

$$\lim_{|t| \to 0} |t|^{-1}[X(\theta + t) - X(\theta) - tV] = 0$$

in p_θ *measure. Then* V *is the derivative in quadratic mean of* X *at* θ *if and only if for every sequence* t_n *tending to a limit* t *the norm* $\varepsilon^{-1}\|X(\theta + \varepsilon t_n) - X(\theta)\|$ *tends to the limit* $\|tV\|$ *as* $\varepsilon \to 0$.

This follows from the fact that convergence in quadratic mean is equivalent to convergence in measure together with convergence of the quadratic norms. Now

$$EX^2(\theta+\varepsilon t_n) = h^2(\theta,\theta+\varepsilon t_n)-\bar{\beta}_\theta(\theta+\varepsilon t_n)$$

where $\bar{\beta}_\theta(t)$ is the mass of p_t, which is p_θ singular.

Thus, if X is differentiable in measure to V it will be differentiable in quadratic mean to V if and only if whenever $t_n \to t$ one has

$$\lim_{\varepsilon\to 0} \varepsilon^{-2}[h^2(\theta,\theta+\varepsilon t_n)-\bar{\beta}_\theta(\theta+\varepsilon t_n)] = E_\theta\,|tV|^2.$$

From this one obtains immediately the following lemma, in which a unit vector t belongs to the closed cone tangent to Θ at θ if there is a sequence v_n such that $\theta+v_n\in\Theta$ and $|v_n|^{-1}v_n \to t$.

LEMMA 8. *Assume that the contiguity condition* (D1) *is satisfied and that Condition* (C2) *is also satisfied for a matrix* $\Gamma(\theta)$.

Assume in addition that at $s = \theta$ *the process* $s \rightleftarrows X(s)$ *is differentiable in* p_θ *measure to a vector* V. *Then the process* $s \rightleftarrows \xi(s)$ *is differentiable in quadratic mean at* $s = \theta$ *if and only if for every unit vector* t *in the closed cone tangent to* Θ *at* θ *one has* $t[\Gamma(\theta)-\Gamma_1]t^T = 0$, *the matrix* Γ_1 *being defined by* $\Gamma_1 = EVV^T$.

COROLLARY. *Suppose that condition* (A1) *is satisfied and that in addition* $C(\theta) = M(\theta)$. *Then the process* $s \rightleftarrows \xi(s)$ *is differentiable in quadratic mean at* $s = \theta$.

PROOF. We have already noted that (A1) combined with the equality $C(\theta) = M(\theta)$ implies all the conditions (D1) to (D4); hence also (C2) for a matrix $\Gamma(\theta)$ equal to $\frac{1}{4}C(\theta)$. The equality $C(\theta) = E\varphi\varphi^T$ is then nothing else than the equality $\Gamma(\theta) = EVV^T$ for $V = \frac{1}{2}\varphi$.

To terminate this section we shall describe a consequence of the stronger assumption called (A2) above. This result implies many of the usual assertions concerning maximum likelihood estimates $\hat{\theta}_n$. In particular, it implies the assertions on the asymptotic behavior of $n^{\frac{1}{2}}(\hat{\theta}_n-\theta)$ made in [1].

As before we shall write $\Delta_n = n^{-\frac{1}{2}}\sum_{j\leqq n}\varphi(x_j, \theta)$ and

$$\Lambda_n(\theta, t) = \sum_{j\leqq n}[\Phi(x_j, t)-\Phi(x_j, \theta)]$$

for variables x_j distributed independently according to p_θ.

PROPOSITION 3. *Suppose that Assumption* (A2) *is satisfied. Let* $S_n(t)$ *be the difference*

$$S_n(t) = \Lambda_n(\theta,\theta+t)-n^{\frac{1}{2}}t\Delta_n+\tfrac{1}{2}ntC(\theta)t^T.$$

There is a function $t \rightleftarrows \eta(|t|)$ *tending to zero as* $|t| \to 0$ *such that for every* $\varepsilon > 0$.

$$P_\theta\{|S_n(t)| \leqq n[\varepsilon+\eta(|t|)]|t|^2 \quad \textit{for all} \quad t \quad \textit{with} \quad \theta+t\in\Theta\}$$

tends to unity as $n \to \infty$.

REMARK. Unless the process $t \rightleftarrows S_n(t)$ is separable or similarly restricted one cannot assert that the set in the probability bracket is measurable. The assertion is then that its *interior* measure tends to unity.

PROOF. One can write the difference $S_n(t)$ in the form

$$n^{-1}S_n(t) = n^{-1}\sum_{j\leq n} t[B(x_j,\theta)-C(\theta)]t^T - n^{-1}\sum_{j\leq n} t[B(x_j,\theta,t)-B(x_j,\theta)]t^T.$$

Let $H[x, |t|] = \sup_s\{\|B(x_j, \theta, s)-B(x, \theta)\|; |s| \leq |t|\}$. According to Assumption (A2) the expectation $\eta_1(|t|) = E_\theta H[x, |t|]$ tends to zero as $|t| \to 0$. We can bound $n^{-1}|S_n(t)|$ by the expression $|t|^2 n^{-1}\sum_{j\leq n} H[x_j, |t|] + |n^{-1}\sum_{j\leq n} t[B(x_j, \theta)-C(\theta)]t^T|$. By the strong law of large numbers the last term will eventually remain inferior to $\frac{1}{2}\varepsilon|t|^2$. For the same reason the term $n^{-1}\sum_{j\leq n} H[x_j, |t|]$ tends pointwise to $\eta_1(|t|)$. Let then $\eta(|t|) = \lim_{\tau\to 0}\eta_1[|t|+|\tau|]$ on the open set where this is less than an arbitrary number $b \geq 1$. Otherwise, take $\eta(t) = \infty$. The finite part of η has in function of $|t|$ only a finite number of jumps larger than $\varepsilon/4$. Thus one can find values $0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_m$ which divide the range of values where η is finite into intervals where η varies by no more than $\varepsilon/4$. To conclude, it is sufficient to note that $H[x, |t|] \leq H[x, \tau_i]$ if $\tau_{i-1} \leq |t| \leq \tau_i$.

From the proposition one can deduce the following result, in which a set K is called tangent to Θ at θ if the Hausdorff distance of the parts of K and Θ situated in the ball centered at θ of radius ε tends to zero faster than ε as $\varepsilon \to 0$.

PROPOSITION 4. *Suppose that Assumption* (A2) *is satisfied, that the matrix* $C(\theta)$ *is non-singular and that* Θ *has a convex tangent set at* θ. *Then there is an* $\alpha > 0$ *with the following property. Let* $\hat{\theta}_n$ *be any point* $\hat{\theta}_n \in \Theta$ *such that*

$$\sum_{j\leq n}\Phi(x_j, \hat{\theta}_n) \geq \sup_{t\in\Theta}\{\sum_{j\leq n}\Phi(x_j, t); |t-\theta| < \alpha\} - n^{-1}.$$

Let T_n *be any point* $T_n \in \Theta$ *such that*

$$n^{-1} + n^{\frac{1}{2}}(T_n-\theta)\Delta_n - \tfrac{1}{2}n(T_n-\theta)C(\theta)T_n-\theta)^T$$
$$\geq \sup\{n^{\frac{1}{2}}t\Delta_n - \tfrac{1}{2}ntC(\theta)t^T; \theta+t\in\Theta\}.$$

Then $n^{\frac{1}{2}}(T_n-\hat{\theta}_n)$ *tends to zero in* P_θ *probability. Furthermore, in the definition of* T_n *one can substitute instead of* Θ *any set* K *tangent to* Θ *at the point* θ.

PROOF. Choose α small enough so that in the notation of Proposition 3 one has $\eta[|t|]t^2 \leq \frac{1}{8}tC(\theta)t^T$ for $|t| \leq \alpha$. Then eventually $\Lambda_n(\theta, \theta+t) \leq n^{\frac{1}{2}}t\Delta_n - \frac{1}{4}ntC(\theta)t^T$ for all $|t| \leq \alpha$. According to the definition of $\hat{\theta}_n$ this gives

$$n^{\frac{1}{2}}(\hat{\theta}_n-\theta)\Delta_n - \tfrac{1}{4}/n(\hat{\theta}_n-\theta)C(\theta)(\hat{\theta}-\theta)^T \geq -n^{-1}.$$

Equivalently, writing C instead of $C(\theta)$ for simplicity

$$[n^{\frac{1}{2}}(\hat{\theta}_n-\theta)-2C^{-1}\Delta_n{}^T]C[n^{\frac{1}{2}}(\hat{\theta}_n-\theta)-2C^{-1}\Delta_n{}^T]^T \leq 4\Delta_n{}^T C^{-1}\Delta_n + n^{-1}.$$

Since $\mathscr{L}[\Delta_n|\theta]$ converges to the Gaussian distribution $\mathscr{N}(0, M)$ it follows that $n^{\frac{1}{2}}|\hat{\theta}_n-\theta|$ is bounded in probability. The same argument applies to T_n. Thus, leaving aside cases which have probability as small as desired, one can replace Θ by the set $\Theta_n = \{t: t\in\Theta, n^{\frac{1}{2}}|t-\theta| < b\}$ or the corresponding set where K replaces Θ. However, for every $\varepsilon > 0$, according to Proposition 3,

$$|\Lambda_n(\theta, \theta+\tau/n^{\frac{1}{2}}) - \tau\Delta_n + \tfrac{1}{2}\tau C\tau^T|$$

is eventually smaller than $\varepsilon|\tau|^2 b^{-2}$ for all τ such that $|\tau| \leqq b$ and $(\theta+\tau) \in \Theta$. This gives

$$n^{\frac{1}{2}}(\hat{\theta}_n-\theta)\Delta_n - \tfrac{1}{2}n(\hat{\theta}_n-\theta)C(\hat{\theta}_n-\theta)^T \geqq n^{\frac{1}{2}}(T_n-\theta)\Delta_n - \tfrac{1}{2}n(T_n-\theta)C(T_n-\theta)^T - n^{-1} - \varepsilon.$$

The result follows by the usual convexity arguments.

When K is a convex cone the immediate implication of the preceding proposition can be described as follows. Suppose $\theta = 0$ for simplicity. Then $n^{\frac{1}{2}}(\hat{\theta}_n-\theta)$ behaves asymptotically like the projection (= closest point) of $C^{-1}\Delta_n{}^T$ on K for the metric induced by C. This implies the appropriate asymptotic normal behavior if K is a linear space. If K is not a linear space one can, in principle, obtain the limiting behavior of $n^{\frac{1}{2}}(\hat{\theta}_n-\theta)$ by projecting an $\mathcal{N}[0, C^{-1}MC^{-1}]$ vector Z on K. This remains true even in certain other cases where K is *not* convex, but the necessary conditions seem hard to describe in a simple manner.

5. Additional results for one dimensional parameters. We have gathered in this section a number of different results which have in common that either they do not extend to k-dimensions or that we do not know of any natural way of extending them. Let us start by recalling a general argument of Kolmogorov. For the purpose, Θ can be any metric space and $t \rightsquigarrow X(t)$, $t \in \Theta$ any stochastic process indexed by Θ. Suppose that for each integer $m = 0, 1, 2, \cdots$ one is given a finite or countable subset S_m of Θ. Assume also that S_0 is reduced to one point $\{t_0\} = S_0$. Let $S = \bigcup_m S_m$ and let $T_m = \bigcup_j \{S_j; j < m\}$. Let C_m be the convex hull of the set of random variables $\{X(\tau); \tau \in T_m\}$. Finally, let ε_m, $m = 1, 2, \cdots$ be a sequence of numbers $\varepsilon_m > 0$ with $\sum \varepsilon_m = a$. Define a number δ_m by the relation

$$\delta_m = \textstyle\sum_{t \in S_m \setminus T_m} P\{\inf[|x(t) - Y|; Y \in C_m] > \varepsilon_m\}.$$

LEMMA (8). *With the notation just described let $\bar{S}$ be the closure of the set S. Assume that $\{X(t); t \in \bar{S}\}$ is a separable process with separating set S itself. Then*

$$P\{\sup_{t \in \bar{S}} |X(t) - X(t_0)| > \textstyle\sum_{m=1}^{\infty} \varepsilon_m\} \leqq \textstyle\sum_{m=1}^{\infty} \delta_m.$$

PROOF. Since S is a separating set for the process $\{X(t); t \in \bar{S}\}$ it is sufficient to prove the result for S itself instead of $\bar{S}$. We shall prove by induction that

$$(*) \qquad P\{\sup_{t \in T_{m+1}} |X(t) - X(t_0)| > \textstyle\sum_{i=1}^{m} \varepsilon_i\} \leqq \textstyle\sum_{i=1}^{m} \delta_i.$$

If this is true, the desired result follows. Indeed, let $A_m = \{\omega: |X(t) - X(t_0)| \leqq a$ for all $t \in T_m\}$. The sets A_m decrease to a certain set A which is precisely the set where $\sup_{t \in S} |X(t) - X(t_0)| \leqq a$. Thus it is enough to prove (*). This inequality certainly holds for $m = 0$, since $T_1 = S_0 = \{t_0\}$ provided only that sums with empty sets of indices be interpreted as equal to zero. Suppose now that the inequality is valid for $i = 0, 1, 2, \cdots, m-1$ with $m-1 \geqq 0$. For every $t \in S_m \setminus T_m$ and every $Y \in C_m$ one can write

$$|X(t) - X(t_0)| \leqq |X(t) - Y| + |Y - X(t_0)|.$$

Thus, taking a supremum on Y over the last term only, we obtain

$$|X(t) - X(t_0)| \leqq |X(t) - Y| + \sup_{\xi \in T_m} |X(\xi) - X(t_0)|.$$

Since this is now true for every $Y \in C_m$ it follows that

$$|X(t)-X(t_0)| \leqq V_m(t)+\sup_{\xi \in T_m}|X(\xi)-X(t_0)|,$$

with $V_m(t) = \inf\{|X(t)-Y|;\ Y \in C_m\}$. It follows that

$$\begin{aligned} P\{\sup_{t\in S_m \cup T_m}|X(t)-X(t_0)| > \varepsilon_m+\textstyle\sum_{i=1}^{m-1}\varepsilon_i\} & \\ \leqq P\{\sup_{t\in S_m\setminus T_m} V_m(t) > \varepsilon_m\} & \\ +P\{\sup_{t\in T_m}|X(t)-X(t_0)| > \textstyle\sum_{i=1}^{m-1}\varepsilon_i\} & \\ \leqq \delta_m+\textstyle\sum_{i=1}^{m-1}\delta_i. & \end{aligned}$$

This concludes the proof of the lemma.

From this point on we shall always assume that Θ is a subset of the real line and that $t \rightsquigarrow X(t)$ is a separable process with set of indices Θ.

In this case consider the quadratic norms defined by

$$\|X(t)-X(s)\|^2 = E|X(t)-X(s)|^2 = h^2(s,t).$$

Let $\sigma(\theta)$ be the value $\sigma(\theta) \in [0, \infty]$ given by

$$\sigma(\theta) = \limsup_{t\to\theta}|t-\theta|^{-1}h(t,\theta).$$

Condition (l) of Section 3 is satisfied at θ if and only if $\sigma(\theta) < \infty$.

DEFINITION 1. Let S be a subset of Θ. The variation of X on the set S is the supremum of all sums of the type $\sum_j\|X(t_{j+1})-X(t_j)\|$ for finite systems of points $\{t_j\}$ such that $t_j < t_{j+1}$ and $t_j \in S$ for all j.

LEMMA 9. *Let S be a closed subset of the line and let $\{X(t);\ t\in S\}$ be a separable stochastic process which is continuous in quadratic mean. Assume that the set S has an infimum $t_0 \in S$. Then for every $B > 0$*

$$P\{\sup_{t\in S}|X(t)-X(t_0)| > B\} \leqq 24L^2/B^2.$$

where L is the variation of X on S.

PROOF. Assume $t_0 = 0$ for simplicity. Extend X to a continuous function defined on the whole of the half line $[0, \infty)$ by interpolating linearly in each interval contiguous to S and leaving X constant on the right of the supremum of S. Let $\tilde{X}$ be the process so extended. Let $s(\tau)$ be the variation of $X(t)$ on the set $[0, \tau)$. This is a continuous nondecreasing function such that $\sup_\tau s(\tau) = L$. For each integer $m > 0$ let $\{\tau_{m,j};\ j = 0, 1, 2, \cdots, 2^m-1\}$ be numbers such that $0 = \tau_{m,0} < \tau_{m,1} \cdots < \tau_{m,j} < \tau_{m,j+1} < \cdots$ and such that $s[\tau_{m,j}] = j2^{-m}L$. Let S_m be this set of points. Construct the successive sets so that $S_m \subset S_{m+1}$. Use this for the sets of Lemma 8 and apply Chebyshev's inequality. This gives

$$\textstyle\sum_{t\in S_m\setminus T_m} P\{\inf_{\tau\in T_m}|X(t)-X(\tau)| > B\varepsilon_m\} \leqq \dfrac{(2^{-m}L)^2}{B^2{\varepsilon_m}^2}2^m.$$

Hence, if $\varepsilon_m = (1-\beta)\beta^{m-1}$, $\beta^2 = \frac{3}{4}$,

$$P\{\sup|X(t)-X(t_0)| > B\} \leqq L^2/B^2 \sum_{m=1}^{\infty} 2^{-m}\varepsilon_m^{-2} \leqq 24L^2/B^2.$$

This concludes the proof of the lemma.

REMARK 1. The continuity in quadratic mean is not really necessary. However, it does imply that any dense set is a separating set.

REMARK 2. Suppose that S is a finite interval $S = [a, b]$ of the line and that the process X satisfies condition (l) at all points $t \in [a, b]$. Then X admits almost everywhere a derivative in quadratic mean $X'(t)$ with $\|X'(t)\| = \sigma(t)$. However, X need not be the Bochner–Lebesgue integral of its derivative, since σ need not be integrable. In fact, the variation of X on $[a, b]$ is given by $L = \int_a^b \sigma(t)\,dt$, and X' will be integrable on $[a, b]$ if and only if X is of bounded variation on $[a, b]$. For a proof see Gel'fand [4]. Consider the process as map from $[a, b]$ to the Hilbert space of equivalence classes of square integrable variables. If (l) is satisfied for every $t \in [a, b]$, then $t \rightsquigarrow X(t)$ is well determined by $X(a)$ and the derivatives $t \rightsquigarrow X'(t)$ where they exist whether or not L is finite, but the integral is of a Denjoy type. A very useful corollary of Lemma 9 is the following result.

COROLLARY. *Suppose that S is a compact subset of the line and that X_0 is a process satisfying the conditions of Lemma 9. Let X_n be a sequence of separable stochastic processes such that $E|X_n(t)-X_n(s)|^2 = E|X_0(t)-X_0(s)|^2$ for all pairs (s, t) of elements of S, and $E|X_n(t_0)|^2 < b < \infty$ for some $t_0 \in S$. Then the sample functions of the processes X_n are almost surely continuous functions. Furthermore, for each $\varepsilon > 0$ there is a bounded equicontinuous subset K of continuous functions such that $P\{X_n(\cdot) \in K\} > 1-\varepsilon$ for all values of n.*

PROOF. Consider the processes Y_n defined by $Y_n(t) = X_n(t) - X_n(t_0)$. Extend them by interpolation to $\tilde{Y}_n$ as in the proof of Lemma 9. On any interval $[\tau_j, \tau_{j+1})$ on which the variation of X_0 is $s(\tau_{j+1}) - s(\tau_j) \leqq L/M$ one can write

$$P\{\sup_t[|\tilde{Y}_n(t) - \tilde{Y}(\tau_j)|; \tau_j \leqq t < \tau_{j+1}] > \varepsilon\} \geqq 24L^2/(\varepsilon^2 m^2).$$

Therefore, it is possible to cover S by m intervals (τ_{j-1}, τ_{j+1}) such that in each interval the process oscillates by not more than 2ε, except perhaps for cases having a total probability inferior to $24L^2[m\varepsilon^2]^{-1}$. Since this can be made as small as one wishes independently of n, the result follows.

For application to statistical problems, let us consider the situation described by the following assumptions.

(B1) The set Θ is a finite or infinite interval of the line. For each θ suppose given a probability measure p_θ, define $h^2(t, \theta) = H^2(p_t, p_\theta)$ as before and let $X(t)$ be the process defined by $X(t) = (dp_t/dp_\theta)^{\frac{1}{2}} - 1$ as in Sections 3 and 4. Let $Y(t) = X(t) - EX(t)$. Consider independent copies $Y_j(t)$, $j = 1, 2, \cdots$ of the process Y.

(B2) The densities dp_t/dp_θ are selected such that the processes Y_j are separable. (This is always possible, of course).

(B3) The process $t \rightsquigarrow X(t)$ is continuous in quadratic mean.

We shall assume that the logarithms of likelihood ratios $\Lambda_n(\theta, t) = \log dP_{t,n}/dP_{\theta,n}$ are taken equal to $2\sum_{j \leqq n} \log[1+X_j(t)]$.

LEMMA 10. *Assume that conditions* (B1), (B2), (B3) *are satisfied. Let* $Z_n(t) = n^{-\frac{1}{2}}\sum_{j\leqq n} Y_j(t)$. *Let* J *be an interval* $J \subset \Theta$ *such that* $\theta \in J$ *and such that the variation of* Y *on* J *is equal to a finite number* L. *Then the following relations hold except perhaps on a set which has probability inferior to* $24L^2B^{-2}$.

(1) $\sup_{t\in J}|Z_n(t)| \leqq B$.
(2) $\Lambda_n(\theta, t) \leqq -nh^2(\theta, t)+2Bn^{\frac{1}{2}}$ *for all* $t \in J$.
(3) $J \cap \{t; \Lambda_n(\theta, t) \geqq 0\} \subset \{t: n^{\frac{1}{2}}h^2(\theta, t) \leqq 2B\}$.

PROOF. Conditions (B2) and (B3) imply that on the interval J the sample functions of the processes Y_j are continuous. Thus the sample functions of the sum Z_n are also continuous, so that Z_n is a separable process. The first relation is then a consequence of Lemma 9. The second relation follows by writing

$$\begin{aligned}\Lambda_n(\theta,t) &= 2\sum_{j\leqq n}\log[1+X_j(t)] \leqq 2\sum_{j\leqq n}X_j(t)\\ &= -nh^2(\theta,t)+2n^{\frac{1}{2}}Z_n(t)\\ &\leqq -nh^2(\theta,t)+2Bn^{\frac{1}{2}}.\end{aligned}$$

The third relation is an obvious consequence of the second.

As an application of this lemma, consider points $\hat{\theta}_n \in J$ such that $\Lambda_n(\theta, \hat{\theta}_n) \geqq 0$ and $\Lambda_n(\theta, \hat{\theta}_n) \geqq \sup_{t\in J}\Lambda_n(\theta, t)-1/n$. In fact, one can often suppose that $\hat{\theta}_n$ actually maximizes $\Lambda_n(\theta, t)$ for $t \in J$, since the sample functions of $\Lambda_n(\theta, t)$ are continuous at each t where $\Lambda_n(\theta, t) > -\infty$, and therefore they assume a maximum on each compact. The lemma says that a $\hat{\theta}_n$ satisfying the conditions just described is such that with high probability $n^{\frac{1}{2}}h^2(\theta, \hat{\theta}_n) \leqq 2B$. In other words, $n^{\frac{1}{2}}h^2(\theta, \hat{\theta}_n)$ is bounded in probability. The more familiar statements of the type that $n^{\frac{1}{2}}(\hat{\theta}_n-\theta)$ is bounded in probability would correspond more or less to the stronger assertion that $nh^2(\theta, \hat{\theta}_n)$ is bounded in probability. We do not know how to prove this without further assumptions. However, we shall now describe a result of this general character. The assumptions made may appear strange. They are easier to understand if one remembers that the function $t \rightleftarrows s(t)$ used below is just as good a parametrization as t itself. For simplicity we will consider only a one-sided situation; that is, of course, immaterial.

(B4) Let $s(\tau)$ be the variation of Y in the interval $[\theta, \tau)$. This is finite for some $\tau > \theta$. Furthermore, there is a number $\alpha^2 > 0$ such that

$$\alpha^2 = \liminf_{t\to\theta} h^2(\theta,t)/s^2(t).$$

PROPOSITION 5. *Let Assumptions* (B1), (B2), (B3), (B4) *be satisfied. Then there is a* $\delta > 0$ *such that if* $a_n > \theta$ *is taken such that* $s(a_n) = 16n^{-\frac{1}{2}}/(\alpha^2\varepsilon)$

$$\Lambda_n(\theta,t) \leqq -\tfrac{1}{2}n\alpha^2s^2(t)+2n^{\frac{1}{2}}Z_n(a_n)$$

for all $t \in [a_n, \theta+\delta)$ *except for cases whose probability is inferior to* $216\varepsilon^2$.

PROOF. Consider on an interval $(\theta, \theta+\delta)$ the process $t \rightleftarrows Y(t)\beta(t)$ where β is a continuous positive decreasing function. The variation of this process in an interval $[a, b) \subset (\theta, \theta+\delta)$ is easily seen to be inferior to

$$M = \int_a^b \beta(t)\, ds(t) - \int_a^b \|Y(t) - Y(a)\|\, d\beta(t)$$
$$\leqq \int_a^b \beta(t)\, ds(t) - \int_a^b s(t)\, d\beta(t).$$

Take for $\beta(t)$ the function $\beta(t) = s^{-2}(t)$. This gives

$$M \leqq 3\int_a^b s^{-2}(t)\, ds(t) \leqq 3/s(a).$$

It follows from Lemma 9 that

$$P\{|Z_n(t) - Z_n(a)| \leqq Bs^2(t), \text{ for all } t \in [a, b)\} \geqq 1 - 216/(s^2(a)B^2).$$

Proceeding as in Lemma 10 one sees that the relation in the probability braces implies

$$\Lambda_n(\theta, t) \leqq -nh^2(\theta, t) + 2n^{\frac{1}{2}}Z_n(a) + 2Bn^{\frac{1}{2}}s^2(t)$$

for all values of $t \in [a, b)$. According to Assumption (B4) there is a $\delta > 0$ such that s remains finite in the interval $[\theta, \theta+\delta]$, and such that $h^2(\theta, t) \geqq \frac{3}{4}\alpha^2 s^2(t)$ for all $t \in [\theta, \theta+\delta]$. Taking $B = \frac{1}{8}\alpha^2 n^{\frac{1}{2}}$

$$\Lambda_n(\theta, t) \leqq -\tfrac{1}{2}\alpha^2 ns^2(t) + 2n^{\frac{1}{2}}Z_n(a).$$

For this value of B one has also

$$s^2(a)B^2 = \tfrac{1}{64}\alpha^4 ns^2(a).$$

For a given $\varepsilon > 0$ one can take a value a_n such that $\frac{1}{64}\alpha^4 ns^2(a_n) = \varepsilon^{-2}$. That is

$$s(a_n) = 16n^{-\frac{1}{2}}/(\varepsilon\alpha^2).$$

The result follows.

Note. The argument assumes implicitly that $s(\tau) > 0$ for $\tau \in (\theta, \theta+\delta)$. The case where s vanishes in an interval is trivial.

COROLLARY 1. *With the notation and assumptions of Proposition* 5, *assume* $s(\tau) > 0$ *for* $\tau \in (\theta, \theta+\delta)$. *Then*

$$P\{\Lambda_n(\theta, t) < 0 \textit{ for all } t \in [a_n, \theta+\delta)\} \geqq 1 - 232\varepsilon^2.$$

PROOF. The inequality in the probability brackets will hold if $2n^{\frac{1}{2}}Z_n(a_n) < \frac{1}{2}n\alpha^2 s^2(a_n)$ and the relation of Proposition 5 holds. According to Chebyshev's inequality

$$P\{|Z_n(a_n)| \geqq (4\varepsilon)^{-1}s(a_n)\} \leqq 16\varepsilon^2.$$

Thus, eliminating cases which have probability at most $232\varepsilon^2$, one can write

$$\Lambda_n(\theta, t) < \frac{2n^{\frac{1}{2}}}{4\varepsilon}s(a_n) - \tfrac{1}{2}n\alpha^2 s^2(t) < \frac{1}{2}\left\{\frac{n^{\frac{1}{2}}}{\varepsilon}s(a_n) - n\alpha^2 s^2(a_n)\right\}.$$

The quantity in the brackets is equal to

$$\frac{16}{\alpha^2\varepsilon^2}-\alpha^2\left(\frac{16}{\alpha^2\varepsilon}\right)^2 = \frac{16}{\alpha^2\varepsilon^2}(1\text{–}16) < 0.$$

The result follows.

COROLLARY 2. *Suppose that assumptions* (B1) *to* (B4) *are satisfied. Suppose also that* $t \neq \theta$ *implies* $p_t \neq p_\theta$. *Let* $J = [\theta, b)$ *be a finite or infinite interval on which* Y *has bounded variation. Assume also that there is an* $\varepsilon > 0$ *such that* $J \cap \{t\colon h^2(\theta, t) \leqq \varepsilon\}$ *is compact. Then*

(1) *The probability that there is a* $\hat{\theta}_n$ *with values in* J *that* $\Lambda_n(\theta, \hat{\theta}_n) = \sup\{\Lambda_n(\theta, t);\ t \in J\}$ *tends to unity as* $n \to \infty$.

(2) *If* $\theta_n{}^*$ *takes values on* J *and is such that* $\Lambda_n(\theta, \theta_n{}^*) \geqq 0$, *and in particular for* $\theta_n{}^* = \hat{\theta}_n$, *the sequence* $n^{\frac{1}{2}}s(\theta_n{}^*)$ *is bounded in probability.*

PROOF. According to Lemma 10, sequences of the type $n^{\frac{1}{2}}h^2(\theta, \theta_n{}^*)$ are bounded in probability. Thus the probability that $\theta_n{}^*$ be in the compact $K = J \cap \{t;\ h^2(\theta, t) \leqq \varepsilon\}$ will tend to unity. This gives the existence of $\hat{\theta}_n$ as before. If $p_t \neq p_\theta$ for $t \neq \theta$ then $\inf\{h^2(\theta, t);\ t \in K \cap [a, b]\} > 0$ for $\theta < a$. It follows that for any $\delta > 0$ the probability that $\theta_n{}^* \in [\theta, \theta+\delta)$ will tend to unity. The last statement is then a consequence of Proposition 5.

Since the condition called (B4) above may appear somewhat unfamiliar the following simple remark may be of interest.

LEMMA 11. *Suppose that condition* (*l*) *is satisfied on an interval* $(\theta-\delta, \theta+\delta)$, $\delta > 0$; *that is to say, suppose that for* $t \in (\theta-\delta, \theta+\delta)$

$$\limsup_{\tau\to 0} |\tau|^{-1} h[t, t+\tau] = \sigma(t) < \infty.$$

Assume also that θ *is a Lebesgue point of* σ *in the sense that*

$$\lim_{\tau\to 0,\, \tau>0} \tau^{-1} \int_{\theta-\tau}^{\theta+\tau} |\sigma(t)-\sigma(\theta)|\, dt = 0.$$

Finally, assume that the usual process ξ *is differentiable in quadratic mean at* $t = \theta$, *or more generally, that*

$$\liminf_{t\to 0} |t|^{-1} h(\theta, \theta+t) > 0.$$

Then condition (B4) *is satisfied at* θ.

PROOF. The variation of the process ξ in an interval $[\theta, t)$ is given by

$$v(t) = \int_\theta^t \sigma(\tau)\, d\tau.$$

It is easily verifiable that the variation of X or Y in the same interval is smaller than or equal to that of ξ. Thus

$$s(t) \leqq \int_\theta^t \sigma(\tau)\, d\tau \leqq (t-\theta)\sigma(\theta) + \int_\theta^t |\sigma(\tau)-\sigma(\theta)|\, d\tau.$$

This gives

$$\liminf\{h(\theta,\theta+t)s^{-1}(\theta+t)\} \geqq \sigma^{-1}(\theta)\liminf|t|^{-1}h(\theta,\theta+t).$$

The result follows.

This lemma is applicable in particular to processes ξ satisfying condition (l) and of bounded variation. For such processes almost all points of Θ are Lebesgue points of the function σ. If, in addition, ξ is differentiable in quadratic mean at θ then

$$\lim\{h(\theta,\theta+t)s^{-1}(\theta+t)\} = 1.$$

Collecting the results established in the preceding propositions, one sees immediately that the following result holds.

PROPOSITION 6. *Let Θ be an interval of the line such that θ is interior to Θ. Assume that the process ξ has bounded variation Θ and that $K = \{t;\, h^2(\theta, t) \leqq \varepsilon\}$ is compact for some $\varepsilon > 0$. Assume that Conditions* (B1), (B2), (B3) *are satisfied. Furthermore, assume that σ is finite in some interval around θ and that θ is a Lebesgue point of σ. Suppose that $\sigma(\theta) > 0$ and that $t \neq \theta$ implies $p_t \neq p_\theta$. Finally, assume that ξ is differentiable in quadratic mean at θ. Then, with probability tending to unity as $n \to \infty$, there exist measurable maximum likelihood estimates $\hat\theta_n$ and $\mathscr{L}[n^{\frac12}(\hat\theta_n-\theta)\,|\,\theta]$ tends to the normal distribution with expectation zero and variance $\gamma(\theta)/[\sigma^2(\theta)+\gamma(\theta)]^2$ where $\gamma(\theta) = E|X'(\theta)|^2 \leqq \sigma^2(\theta)$.*

PROOF. The proposition asserts measurability of $\hat\theta_n$. This can easily be verified as follows. Divide the compact K covering it by intervals $[a_j, a_{j+1})$ of length 2^{-k}. Number the intervals in their natural order on the line. Take the first j such that $\Lambda_n(\theta, t)$ reaches its maximum in $[a_j, a_{j+1})$. Let T_k be equal to a_j. If the covering of K is a partition which is refined in the ordinary binary manner and the sample functions of Λ_n are continuous, the functions T_k will converge as $k \to \infty$ to a choice of $\hat\theta_n$. To prove the final statement of the proposition, note that with high probability one can restrict oneself to an interval of the type $(\theta - a/n^{\frac12},\, \theta + a/n^{\frac12})$ since the variation $s(t)$ is roughly proportional to $t-\theta$. Consider then the process $\tau \rightsquigarrow W_n(\tau) = \frac12\Lambda_n(\theta,\, \theta+\tau/n^{\frac12})$ in the same interval. The differentiability in quadratic mean implies that $V_n(\tau)$ is asymptotically normal. The corollary of Lemma 9 implies that the convergence to Gaussian distributions occurs in the Prohorov sense for the uniform norm in the space of continuous functions on $[-a, +a]$. It follows by a fairly standard argument that

$$\sup_{|\tau|\leqq a}\left|W_n(\tau) - V_n(\tau) + \tfrac12 \operatorname{Var} V_n(\tau)\right| \to 0$$

in probability as $n \to \infty$. Letting $X_j'(\theta)$ be the derivative of $X_j(t)$ at $t = \theta$, and letting $\Delta_n = n^{-\frac12}\sum_{j\leqq n} X_j'(\theta)$, one can replace in a similar manner the process $V_n(t)$ by $-\frac12 nh^2(\theta,\, \theta+t/^{\frac12}n) + t\Delta_n$, and the variance of $V_n(\tau)$ by $\gamma(\theta)\tau^2$. Thus

$$\sup_{|\tau|\leqq a}\left|W_n(\tau) - \tau\Delta_n + \tfrac12[\sigma^2(\theta)+\gamma(\theta)]\tau^2\right|$$

tends to zero in probability as $n \to \infty$. It follows that $n^{\frac12}(\hat\theta_n - \theta) - [\sigma^2(\theta)+\gamma(\theta)]^{-1}\Delta_n$ tends to zero in probability. Hence the result.

To terminate, let us mention an extension of a lemma of Hájek [5] which offers an often convenient way of checking our condition of differentiability in quadratic mean.

Let $\{\mathfrak{X}, \mathscr{A}, \nu\}$ be a σ-finite measure space. Let Θ be an interval of the line.

DEFINITION. Let g be a real-valued function defined on $\{\mathfrak{X}\times\Theta\}$. We shall say that g is simply absolutely continuous in θ if it is measurable in x for each θ and if there is a function $(x, \theta) \rightsquigarrow \psi(x, \theta)$ such that for every pair (s, t), $s < t$ of elements of Θ the relation

$$g(x, t)-g(x, s) = \int_s^t \psi(x, \tau)\, d\tau$$

holds almost everywhere ν, the integral on the right being for every $x \in \mathfrak{X}$ a Lebesgue integral.

If g is a function satisfying the above definition fix s and consider g^* defined by

$$g^*(x, t) = g(x, s)+\int_s^t \psi(x, \theta)\, d\theta.$$

Then $g^*(x, t) = g(x, t)$ except on a ν null set A_t. Thus g^* is a possible "version" of the family $\{g(x, t); t \in \Theta\}$. This version is continuous in t for each x and measurable in x for each t. Thus it is jointly measurable. It follows that

$$\psi^*(x, \theta) = \lim\sup_{\tau\to 0} \tau^{-1}[g^*(x, \theta+\tau)-g^*(x, \theta)]$$

is also jointly measurable.

According to Lebesgue one must have $\psi^*(x, t) = \psi(x, t)$ except perhaps on a set B_x of values of t which has Lebesgue measure zero. It follows that

$$\int |\psi^*(x, t)-\psi(x, t)|\, dt = 0$$

and that one may as well assume that ψ itself is selected jointly measurable.

PROPOSITION 7. *Assume that Θ is an interval of the line and that $\{\mathfrak{X}, \mathscr{A}, \nu\}$ is a σ-finite measure space. Let g be a function defined on $\mathfrak{X}\times\Theta$. Suppose that g is simply absolutely continuous for a function ψ which is jointly measurable.*

Let $\sigma^2(t) = \int \psi^2(x, t)\nu(dx)$ and assume that θ is a Lebesgue point of the map $t \rightsquigarrow \sigma(t)$. Let G_t be the equivalence class of $x \rightsquigarrow g(x, t)-g(x, \theta)$, then the map $t \rightsquigarrow G_t$ admits the class of $x \rightsquigarrow \psi(x, \theta)$ as a derivative in quadratic mean (for ν) at $t = \theta$ if and only if at the point θ the function g admits ψ for derivative in ν-measure on each set $A \in \mathscr{A}$ of finite ν-measure.

PROOF. The necessity of the condition is clear. To prove the sufficiency note that for $t > 0$ one can write

$$\begin{aligned}
\|t^{-1}G_t\|^2 &= \int t^{-2}\,|g(x, \theta+t)-g(x, \theta)|^2\nu(dx)\\
&= \int |t^{-1}\int_\theta^{\theta+t}\psi(x, \tau)\, d\tau|^2\nu(dx)\\
&\leqq \int \{t^{-1}\int_\theta^{\theta+t}|\psi(x, u)|\, du\}\{t^{-1}\int_\theta^{\theta+t}|\psi(x, \tau)|\, d\tau\}\nu(dx)\\
&\leqq t^{-2}\int_\theta^{\theta+t}\int_\theta^{\theta+t}\{\int |\psi(x, u)|\, \psi(x, \tau)|\nu(dx)\}\, du\, d\tau\\
&\leqq t^{-2}\int_\theta^{\theta+t}\int_\theta^{\theta+t}\sigma(u)\sigma(\tau)\, du\, d\tau = |t^{-1}\int_\theta^{\theta+t}\sigma(\tau)\, d\tau|^2.
\end{aligned}$$

Thus

$$||t^{-1}G_t|| \leqq \sigma(\theta)+t^{-1}\int_\theta^{\theta+t}|\sigma(u)-\sigma(\theta)|\,du.$$

It follows then from Fatou's lemma that

$$\lim_{t\to 0}||t^{-1}G_t|| = \sigma(\theta).$$

The desired conclusion is then a consequence of the usual argument in Hilbert space.

REMARK 1. The proposition may be applied to the case where ν is Lebesgue measure on the line and $g(x, \theta) = g(x-\theta)$. In this case $\sigma(t)$ is constant in t and the result reduces to Hájek's statement.

REMARK 2. In the application to statistical problems the function $g(x, \theta)$ will be the square root of a probability density $f(x, \theta)$. In this case if the set of points x at which $f(x, \theta) = 0$ has ν-measure zero, then differentiability in measure of f implies differentiability in measure of g at $t = \theta$. Furthermore, simple absolute continuity of g implies that of f, but the *converse is not true* as simple examples will show. Because of this it is perhaps convenient to note the following. Suppose that $\lim_{\tau\to 0}\tau^{-1}[g(x, t+\tau)-g(x, t)] = \psi(x, t)$ except on a set A_t such that $\nu(A_t) = 0$ and that ψ is jointly measurable. Furthermore, suppose that θ is a Lebesgue point of $t \rightsquigarrow \sigma(t) = [\int\psi^2(x, t)\nu(dx)]^{\frac{1}{2}}$. When ν is a finite measure the absolute continuity of g around θ will be assured if one knows in addition that

$$\lim_{\tau\to 0}\tau^{-1}|g(x,t+\tau)-g(x,t)| < \infty,$$

except perhaps on a countable set of values of t which may depend on x.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[2] DOOB, J. L. (1936). Statistical estimation. *Trans. Amer. Math. Soc.* **39** 410–421.

[3] DUGUÉ, D. (1937). Application des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation. *J. École Polytechnique* **3** 305–374.

[4] GEL'FAND, I. M. (1938). Abstrakte funktionen und lineare operatoren. *Mat. Sb.* **4(46)** 238–286.

[5] HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

[6] KAKUTANI, S. (1948). On equivalence of infinite product measures. *Ann. of Math.* **49** 214–224.

[7] LECAM, L. (1966). Likelihood functions for large numbers of independent observations in *Festschrift for J. Neyman*, ed. F. N. David. Wiley, New York. 167–187.

[8] LECAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 37–98.

[9] SAKS, S. (1937). *Theory of the Integral.* Monografie Matematyczne, Warsaw. Dover, New York (1964).

[10] SHEPP, L. (1965). Distinguishing a sequence of random variables from a translate of itself. *Ann. Math. Statist.* **36** 1107–1112.

[11] WALD, A. (1942). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

[12] WILKS, S. S. (1938). Shortest average confidence intervals from large samples. *Ann. Math. Statist.* **9** 166–175.