ON THE ASYMPTOTIC THEORY OF ESTIMATION AND TESTING HYPOTHESES

L. LE CAM
UNIVERSITY OF CALIFORNIA

1. Introduction

The purpose of the present paper is to develop a systematic method of reduction of problems of estimation and testing hypotheses to similar problems on normal distributions. The method proposed is only valid asymptotically and but little is known about its performance for samples of finite size.

To describe more precisely the questions investigated it would be necessary to proceed to an historical review of large sample theory. For brevity and simplicity we shall restrict ourselves to the key papers of J. Neyman and A. Wald. The first of these authors introduced in [1] what are called best asymptotically normal regular estimates (B.A.N. for short). The situation considered by Neyman is one in which the distributions are of a multinomial nature, but the same techniques apply to families of the Koopman-Darmois type (see [2]). Part of the motivation for the introduction of B.A.N. estimates is that the maximum likelihood estimates are, even in such a "simple" case, very often difficult to obtain. Furthermore, the B.A.N. estimates behave asymptotically very much like the maximum likelihood estimates.

In sharp contrast with the preceding, Wald [3] considers classes of densities restricted only by regularity conditions. In such a case, sufficient statistics of fixed dimensionality do not usually exist, so that the methods used by Wald are by necessity different from those of Neyman. Wald confines his attention to maximum likelihood estimates and tests based on these estimates. A fundamental result of Wald is that, under certain conditions, the maximum likelihood estimates are "asymptotically sufficient." Further, by means of suitable *set* transformations it is possible to associate to each test problem on the original distributions a closely related, though not equivalent, problem on normal distributions. The asymptotic sufficiency of the maximum likelihood estimates would make Neyman's techniques available to the statistician if only he could obtain the values of the estimates. Since Wald's reasoning relies heavily on the fact that maximum likelihood estimates are roots of the corresponding equations, it is not at all clear that the same results would remain valid for approximate maximum likelihood estimates. It is even less clear that the results would hold when the maximum likelihood estimates are not solutions of the relevant equations, a circumstance which occurs often on boundaries of the parameter space.

This paper was prepared with the partial support of the Office of Naval Research, and of the Office of Ordnance Research, U.S. Army, under Contract DA-04-200-ORD-355.

In the present paper the author has attempted to describe a method by which results similar to those of Wald's can be obtained, still retaining a large amount of freedom in the choice of the estimates. It will be seen that even though maximum likelihood estimates might not be consistent or even might not exist, it is possible under certain conditions to obtain estimates which are consistent, asymptotically normal and asymptotically sufficient by using the averages of the logarithmic derivatives of the densities. Although the existence of such estimates is proved by a procedure which provides a particular estimate, it is clear that in most cases a whole class of such estimates is readily available.

Finally, the method presented here leads to proofs of asymptotic optimality of certain tests proposed by Neyman in [4].

2. Notation and assumptions

Although the following assumptions are rather stringent, they seem to cover a large variety of problems. They will be referred to as assumptions (A) and used throughout unless the contrary is explicitly specified.

Let Θ be a parameter set indexing the states of nature. Let $\mathcal{X}$ be the space where the observable variables take their values. It is assumed that a σ-field $\mathcal{A}$ of subsets of $\mathcal{X}$ has been chosen and that to each $\theta \in \Theta$ there corresponds a probability measure P_θ on $\mathcal{A}$.

ASSUMPTION 1. *$\mathcal{X}$ is a Euclidean space and $\mathcal{A}$ is the field of Borel subsets of $\mathcal{X}$. The space Θ is a locally compact subset of an r-dimensional Euclidean space $\mathcal{E}$.*

ASSUMPTION 2. *Observations can be made on a sequence of independent random variables $\{X_j\}$, $j = 1, 2, \cdots$, taking their values in $\mathcal{X}$. The distribution of the sequence $\{X_j\}$ is the product measure corresponding to one of the measures P_θ, $\theta \in \Theta$.*

ASSUMPTION 3. *$P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$.*

ASSUMPTION 4. *There exists on $\mathcal{A}$ a measure ν, finite on compacts and such that for every $\theta \in \Theta$ the measure P_θ admits a density with respect to ν.*

It will be assumed that for each $\theta \in \Theta$ a particular value $p(x, \theta)$ of the density $dP_\theta/d\nu$ has been selected.

ASSUMPTION 5. *The function $\Phi(x, \theta) = \log p(x, \theta)$ is for each $x \in \mathcal{X}$ a finite continuous function of θ.*

We pass now to assumptions relative to the differentiability of Φ. It happens to be particularly convenient to make use of a rather specific form of the remainder term in Taylor's formula, and the assumptions are stated with due consideration of this. It is often convenient and sometimes necessary to make use of local coordinate systems instead of a fixed Cartesian system as assumed here. The modifications necessary for such considerations would complicate the arguments without bringing in essentially new features. Furthermore, the necessary modifications are rather obvious so that we shall limit ourselves to the simplest case.

It will be assumed that Θ does not have any isolated points and further that there is a set Ω containing Θ such that each point of Θ possesses in Ω a convex neighborhood.

Furthermore, it will be assumed that to each couple (x, t) with $x \in \mathcal{X}$ and $t \in \Omega$ there corresponds a $1 \times r$ matrix $A(x, t)$ and an $r \times r$ symmetric matrix $B(x, t)$ in such a way that the following properties hold.

ASSUMPTION 6. *For each $x \in \mathcal{X}$ the functions $t \to A(x, t)$ and $t \to B(x, t)$ are continuous functions of t on Ω. For each $t \in \Omega$ the functions $x \to A(x, t)$ and $x \to B(x, t)$ are Borel measurable in x.*

ASSUMPTION 7. *For every $\theta \in \Theta$ there is a neighborhood V_θ of θ in Ω and a numerical function $H_\theta(x)$ such that*

(a) *For a given $\theta \in \Theta$ the integrals*

$$(1) \qquad \int H_\theta(x)\, p(x, t)\, d\nu$$

are uniformly convergent for $t \in V_\theta \cap \Theta$.

(b) *If $t \in V_\theta$ every element of $B(x, t)$ is bounded in absolute value by $H_\theta(x)$.*

(c) *If t and τ are two elements of $V_\theta \cap \Theta$ then*

$$(2) \qquad \Phi(x, t) - \Phi(x, \tau) = A(x, \tau)(t-\tau) - \tfrac{1}{2}(t-\tau)'B(x; \tau, t)(t-\tau)$$

with

$$(3) \qquad B(x; \tau, t) = 2\int_0^1 (1-\lambda) B[x; \tau + \lambda(t-\tau)]\, d\lambda.$$

ASSUMPTION 8. *For every $\theta \in \Theta$ we have*

$$(4) \qquad E\{A(X, \theta) \mid \theta\} = 0$$

$$(5) \qquad E\{B(X, \theta) \mid \theta\} = E\{A'(X, \theta) A(X, \theta) \mid \theta\} = \Gamma(\theta),$$

say. Furthermore, $\Gamma(\theta)$ is positive definite.

A point $\{x_1, x_2, \cdots, x_n\}$ in the product of n copies of $\mathcal{X}$ will be denoted by z_n and the corresponding random variable by Z_n. If the distribution of Z_n is the product measure P_θ^n the corresponding density with respect to the product measure ν^n will be denoted by $\prod_{j=1}^n p(x_j, \theta) = p_n(z_n, \theta)$.

To simplify formulas we shall denote by $A_n(t)$ the row matrix $A_n(t) = (1/n)\sum_{j=1}^n A(x_j, t)$ and by $B_n(t)$ the square matrix $B_n(t) = (1/n)\sum_{j=1}^n B(x_j, t)$. Similarly $\Phi_n(z_n, t)$ will be used to represent $\sum_{j=1}^n \Phi(x_j, t)$. It will often be convenient to consider $A_n(t)$ and $B_n(t)$ as random variables and this without changing the notation. The interpretation will always be clear from the context. Norms of vectors or matrices will be denoted by double bars.

Let S be a Euclidean space and let $\mathcal{M}$ be the space of bounded signed measures on S. On $\mathcal{M}$ we shall consider a norm defined by $\|\mu\| = \mu^+(S) + \mu^-(S) = \sup\{\mu(A) + \mu(A^c);\ A \in \mathcal{A}\}$. Furthermore, we shall use on $\mathcal{M}$ the topology $\mathcal{T}_c$ and the associated uniform structure defined by neighborhoods of the origin of the generic type

$$(6) \qquad V = \{\mu : |\int u_j\, d\mu| \leqq 1; \quad j = 1, 2, \cdots, k\},$$

where the functions u_j are continuous and bounded on S. The corresponding structures obtained by restricting the u_j's to be continuous and to vanish outside a compact will be called the vague topology and the vague uniform structure.

For subsets of a vector space the symbols $A + B$ and ϵA with ϵ real will be given the usual meaning.

If $\mathcal{X}_k$ is the product of k copies of $\mathcal{X}$, an estimate depending on z_k is a function T_k from $\mathcal{X}_k$ to the Euclidean space $\mathcal{E}$ containing Θ. Such estimates will be called strict if their range is a subset of Θ.

We shall not assume that estimates are necessarily measurable functions, the reason for this being that the measurability of estimates obtained by the usual procedures is often awkward to check. If measurability is used or needed we shall explicitly mention the fact.

The reader should beware of the fact that lack of measurability entails some complications. Thus, almost sure convergence does not imply convergence in probability. Further, the theorems due to Slutsky [5] and Cramér [6] have to be reinterpreted since nonmeasurable functions do not have distributions. The necessary modifications being evident, we shall not dwell on this at any length. Upper and lower asterisks will be used occasionally for upper integrals, outer measures, etc.

The assumptions (A) given above being somewhat stringent, it is useful to keep in mind examples where they are satisfied. Note that the assumptions (A) do not imply that Θ has any interior point in $\mathcal{E}$. In fact, Θ could be nowhere dense in $\mathcal{E}$.

Some extreme examples satisfying assumptions (A) have been used by C. Kraft and the author [7] to show that assumptions (A) do not imply the consistency of maximum likelihood estimates, or when the method applies, the consistency of minimum χ^2 estimates.

More specifically, assumptions (A) do not imply the existence of the maximum likelihood estimates. Even if the maximum likelihood estimates exist and are uniquely defined, assumptions (A) do not imply their consistency. Examples of a very regular nature satisfying assumptions (A) are the exponential families constructed as follows. Let μ be an arbitrary measure on a Euclidean space $\{\mathcal{X}, \mathcal{A}\}$. The set S of values of $s \in \mathcal{X}$ for which the integral $K^{-1}(s) = \int \exp(s'x)d\mu$ is finite is always a convex subset of $\mathcal{X}$. If the linear dimension of S is smaller than the dimension of $\mathcal{X}$, it is possible to reduce the dimensionality of the variable X by the same amount; hence we shall assume that $\mathcal{X}$ and S have both been reduced to their minimum dimension. Let s_0 be a point interior to S in $\mathcal{X}$ and let ν be the measure defined by $\nu(A) = \int_A \exp(s_0'x)d\mu$. Let $p(x, s) = K(s) \exp [x'(s - s_0)]$ be considered as a density with respect to ν. If Θ is an arbitrary open subset of S the family $\{p(x, s), s \in \Theta\}$ satisfies assumptions (A). More generally, let $\tilde{\Theta}$ be an arbitrary open subset of a Euclidean space. Let $\theta \to s(\theta)$ be a twice continuously differentiable map from $\tilde{\Theta}$ to the interior of S such that $\theta_1 \neq \theta_2$ implies $s(\theta_1) \neq s(\theta_2)$. The resulting family $\{p[x, s(\theta)]; \theta \in \tilde{\Theta}\}$ satisfies assumptions (A). The same assumptions (A) are still satisfied if $\tilde{\Theta}$ and $s(\theta)$ being as above, one considers only the family $\{p[x, s(\theta)]; \theta \in \Theta\}$ where Θ is a subset of $\tilde{\Theta}$, locally compact and without isolated points. As an extreme example of circumstances where assumptions (A) are satisfied consider the following: Let Ω be an open subset of the real line subject only to the restriction that its connected components have finite length. For instance, Ω might be the complement in $(-1, +2)$ of the Cantor set, or Ω might be the real line deprived of the integers. Order the intervals composing Ω in a sequence $\{J_k\}$, $k = 1, 2, \cdots$, and order the rationals of the open interval $(1, 2)$ in a sequence $\{a_k\}$, $k = 1, 2, \cdots$. For $\theta \in J_k$ let $\rho(\theta) = a_k$, let $\lambda(\theta)$ be the reciprocal of the length of J_k and let $\beta(\theta)$ be the lower bound of the interval J_k.

Let X be a two-dimensional random vector having a normal distribution with the identity as covariance matrix. Suppose that for $\theta \in \Omega$ the expectation of the coordinates of X are, respectively, $\rho(\theta) \cos \{2\pi\lambda(\theta)[\theta - \beta(\theta)]\}$ and $\rho(\theta) \sin \{2\pi\lambda(\theta)[\theta - \beta(\theta)]\}$. One verifies immediately that one can satisfy assumptions (A) by taking a sequence of independent random vectors having the above characteristics, choosing Θ to be either Ω itself or any locally compact subset of Ω without isolated points.

3. Preliminary lemmas

In order to make the proofs of the main theorems more understandable and yet reasonably complete, some of the relevant lemmas have been collected in this section.

Let $B_n(t) = (1/n) \sum_{j=1}^{n} B(X_j, t)$ and let $C(t, \tau) = E\{B(X, t) | \tau\}$.

LEMMA 1. *Let assumptions (A) be satisfied. Every $\theta \in \Theta$ possesses a neighborhood U in Ω such that for every positive ϵ one can find an integer N depending on ϵ and θ only for which*

$$\sup_{\tau \in U \cap \Theta} P^* \{ \sup_{n \geq N} \sup_{t \in U} \| B_n(t) - C(t, \tau) \| > \epsilon | \tau \} < \epsilon . \tag{7}$$

PROOF. Let V_θ be the neighborhood described in assumption 7 and let $H(x)$ be the corresponding function. Let U be a compact neighborhood of θ contained in V_θ and such that $U \cap \Theta$ be compact. Let δ be a positive number. Denote by C the space of all continuous numerical functions on U considered as a Banach space for the norm $\|g\| = \sup \{ |g(t)| ; t \in U \}$. Let $f(x, t)$ be an element of the matrix $B(x, t)$ and let $\{g_k\}$ be a denumerable dense set in C.

For every integer k let S'_k be the set of values of x such that $\|g_k(t) - f(x, t)\| < \delta$. Clearly S'_k is measurable. The sequence $\{S'_k\}$ can be disjointed, giving a sequence $\{S_k\}$ defined by $S_1 = S'_1$ and $S_{k+1} = S'_{k+1} \cap (\bigcup_{j \leq k} S_j)^c$.

Let

$$h(x, t) = \sum_{k=1}^{\infty} I_{S_k}(x) g_k(t) . \tag{8}$$

By construction

$$\sup_x \sup_t | h(x, t) - f(x, t) | \leq \delta \tag{9}$$

and

$$| h(x, t) | \leq H(x) + \delta . \tag{10}$$

Let E_k be the set $E_k = \bigcup_{j>k} S_j$. The compactness of $U \cap \Theta$ and the continuity of the map $\tau \rightarrow p(x, \tau)$ implies that as k tends to infinity the integrals

$$\int I_{E_k}(x) [H(x) + \delta] p(x, \tau) d\nu \tag{11}$$

tend to zero uniformly for $\tau \in U \cap \Theta$. Hence there exists m so large that

$$\int I_{E_m}(x) [H(x) + \delta] p(x, \tau) d\nu < \frac{\delta\epsilon}{2} \tag{12}$$

for every $\tau \in U \cap \Theta$.
Let

$$\begin{cases} h'(x, t) = \sum_{j=1}^{m} I_{S_j}(x) g_j(t) \\ h''(x, t) = \sum_{j>m} I_{S_j}(x) g_j(t) . \end{cases} \tag{13}$$

Clearly

(14) $$\sup_t |h''(x, t)| = K(x) \leqq H(x) + \delta .$$

Hence

(15) $$\sup_{\tau \in U \cap \Theta} \int_{E_m}^* K(x) p(x, \tau)\, d\nu \leqq \frac{\delta\epsilon}{2} .$$

Let M be a bound for the norms $\|g_j\|, j = 1, 2, \cdots, m$. Then according to the martingale inequalities

(16) $$P_\tau^* \left\{ \sup_{n \geqq N} \sup_{t \in U} \left| \frac{1}{n} \sum_{i=1}^n I_{S_j}(X_i)\, g_j(t) - E[I_{S_j}(X)\, g_j(t) \mid \tau] \right| > \delta \right\} \leqq \frac{M^2}{N\delta^2} .$$

From this one concludes that

(17) $$\sup_{\tau \in U \cap \Theta} P_\tau^* \left\{ \sup_{n \geqq N} \sup_{t \in U} \left| \frac{1}{n} \sum_{i=1}^n h'(X_i, t) - E[h'(X, t) \mid \tau] \right| > \delta \right\} \leqq \frac{m^3 M^2}{N\delta^2} .$$

Furthermore,

(18) $$P_\tau^* \left\{ \sup_{n \geqq N} \frac{1}{n} \sum_{i=1}^n K(X_j) \geqq \delta \right\} \leqq \frac{\epsilon}{2} ,$$

thus choosing $\delta = \epsilon/3$ and then $N = 18m^3M^2/\epsilon^2$ gives

(19) $$\sup_{\tau \in U \cap \Theta} P_\tau^* \left\{ \sup_{n \geqq N} \sup_{t \in U} \left| \frac{1}{n} \sum_{i=1}^n f(X_i, t) - E[f(X_i, t) \mid \tau] \right| > \epsilon \right\} < \epsilon .$$

Repeating the argument for each element of $B(x, t)$ gives the desired result.

LEMMA 2. *Let assumptions (A) be satisfied. Then the integrals*

(20) $$\int A'(x, \theta)\, A(x, \theta)\, p(x, \theta)\, d\nu$$

are uniformly convergent on compact subsets of Θ. *Consequently*

(21) $$\mathcal{L} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n A'(X_j, \theta) \mid \theta \right\} \to \mathcal{N}[0, \Gamma(\theta)]$$

in the ordinary ($\mathcal{T}_c$) *sense uniformly on the compacts of* Θ.

PROOF. Let K be a compact of Θ and let θ_0 be a point of K. Slutsky's theorems and the continuity of $p(x, \theta)$ with respect to θ imply that

(22) $$\| \mathcal{L}\{ A'(X, \theta_0) \mid \theta_n \} - \mathcal{L}\{ A'(X, \theta_0) \mid \theta_0 \} \|$$

tends to zero if θ_n tends to θ_0.

Since $A'(x, \theta_n) = A'(x, \theta_0) - M(x; \theta_0, \theta_n)(\theta_n - \theta_0)$ for some average $M(x; \theta_0, \theta_n)$ of $B(x, t)$ between θ_0 and θ_n, Slutsky's theorems also imply that

(23) $$\mathcal{L}\{ A'(X, \theta_n) \mid \theta_n \} \to \mathcal{L}\{ A'(X, \theta_0) \mid \theta_0 \}$$

in the $\mathcal{T}_c$ sense. Furthermore, $\Gamma(\theta) = E\{A'(X, \theta)A(X, \theta) | \theta\}$ is a continuous function of θ. Let α be an arbitrary fixed vector. Let $Y_\theta = A(X, \theta)\alpha$ and let $F_\theta = \mathcal{L}\{Y_\theta | \theta\}$. Let μ_θ be the measure defined by $\mu_\theta(S) = \int_S y^2 dF_\theta$. We have just seen that if $\theta_n \to \theta_0$

then $F_{\theta_n} \to F_{\theta_0}$ and $\int y^2\, dF_{\theta_n} \to \int y^2\, dF_{\theta_0}$. Therefore, $\mu_{\theta_n} \to \mu_{\theta_0}$ for $\mathcal{T}_c$. Hence for every $\epsilon > 0$ there is a number b such that

$$(24) \qquad \int_{|y|>b} y^2 dF_{\theta_n} < \epsilon$$

for every integer n. This, repeated for an orthonormal system of values of α, implies the desired result. Indeed if the integrals were not uniformly convergent on K there would exist a sequence $\{\theta_n\}$ tending to a $\theta_0 \in K$ for which the above would not hold.

The statement on the convergence to normal distributions follows immediately [8].

Our next lemma is relative to modifications of normal densities. Let the set Θ and the matrix $\Gamma(\theta)$ be as in the assumptions (A). Let λ denote the Lebesgue measure on the Euclidean space $\mathcal{E}$. Further, let

$$(25) \qquad g[t;\ \tau,\ \theta] = \frac{[\det \Gamma(\theta)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(t-\tau)'\Gamma(\theta)(t-\tau)\},$$

for every $\theta \in \Theta$ and every couple of points t and τ of $\mathcal{E}$. Let f be a not necessarily measurable function from $\mathcal{E}$ to $\mathcal{E}$. Let $\tilde{g}$ be the function obtained by substituting $f(t)$ for t in $g(t;\ \tau,\ \theta)$.

LEMMA 3. *For every compact $C \subset \Theta$ and every $\epsilon > 0$ there exists a $\delta > 0$ such that if* $\sup_t |f(t) - t| < \delta$ *then*

$$(26) \qquad \int^* |\tilde{g}(t;\ \tau,\ \theta) - g(t;\ \tau,\ \theta)|\, d\lambda < \epsilon$$

for every $\tau \in \mathcal{E}$ and every $\theta \in C$.

PROOF. Since g is always positive we can write

$$(27) \qquad \int^* |\tilde{g} - g|\, d\lambda = \int^* g \left|\frac{\tilde{g}}{g} - 1\right| d\lambda$$
$$= \int^* g\, |\exp\{-\tfrac{1}{2}[f(t) - \tau]'\Gamma(\theta)[f(t) - \tau] + \tfrac{1}{2}(t-\tau)'\Gamma(\theta)(t-\tau)\} - 1|\, d\lambda .$$

Letting $\psi(t) = f(t) - t$, this becomes

$$(28) \qquad \int^* |\tilde{g} - g|\, d\lambda = \int^* g\, |\exp\{-\tfrac{1}{2}\psi'\Gamma\psi - \psi'\Gamma(t-\tau)\} - 1|\, d\lambda;$$

therefore,

$$(29) \qquad \int^* |\tilde{g} - g|\, d\lambda$$
$$\leqq \int^* g\, |\exp\{+\tfrac{1}{2}\psi'\Gamma\psi + [\psi'\Gamma\psi]^{1/2}[(t-\tau)'\Gamma(t-\tau)]^{1/2}\} - 1|\, d\lambda .$$

Let β^2 be a bound for $\psi'\Gamma\psi$ when t runs through $\mathcal{E}$ and θ through C, let $t - \tau = [\Gamma(\theta)]^{-1/2}\xi$ and let $g_1(\xi)$ be the density of an $\mathcal{N}(0, I)$ variable. This gives

$$(30) \qquad \int^* |\tilde{g} - g|\, d\lambda \leqq \int g_1(\xi)\,[e^{\beta^2/2 + \beta\|\xi\|} - 1]\, d\lambda$$

and implies the desired result.

4. Existence of consistent estimates

In the next sections frequent use will be made of estimates T_n which not only are consistent but are such that $\sqrt{n}\|T_n - \theta\|$ is "bounded in probability." We have already mentioned that under assumptions (A) maximum likelihood estimates need not be consistent. Thus, the question arises whether there exist consistent estimates of any nature. This is implied by the following general lemma.

LEMMA 4. *Let assumptions* (1), (2), *and* (3) *be satisfied. Furthermore, assume that the map* $\theta \to P_\theta$ *is continuous in the sense that if* θ_n *tends to* θ_0 *then* P_{θ_n} *tends to* P_{θ_0} *in the ordinary* $\mathcal{T}_c$ *sense.*

Then there exists a sequence $\{T_n\}$ *of measurable strict estimates such that for every positive* ϵ *and every compact subset* K *of* Θ *the quantity*

$$\sup_{\theta \in K} P\{\|T_n - \theta\| > \epsilon \mid \theta\} \tag{31}$$

tends to zero as n tends to infinity.

PROOF. In the space $\mathcal{M}$ of bounded signed measures on $\{\mathcal{X}, \mathcal{A}\}$ let $V = \{\mu : |\int u_j d\mu| \leqq 1 \text{ for } j = 1, 2, \cdots, k\}$ be a typical neighborhood of the origin. Let $\mathcal{M}_1$ be the subset of $\mathcal{M}$ constituted by probability measures. Let F be an arbitrary element of $\mathcal{M}_1$ and for a sample point $z = \{x_1, x_2, \cdots, x_n\}$ let $F(n, z)$ be the corresponding empirical distribution. If M is a bound for the functions u_j, one can write

$$P\{F(n, Z) \in F + V \quad \text{for every} \quad n \geqq N \mid F\} \geqq 1 - \frac{kM^2}{N}. \tag{32}$$

Let $\{K_j\}$, $j = 1, 2, \cdots$, be a sequence of compact subsets of Θ such that K_{j+1} is the closure of a set open in Θ and containing K_j and such that $\bigcup_j K_j = \Theta$. Let $S_j = \{P_\theta; \theta \in K_j\}$. Since the map $\theta \to P_\theta$ is assumed to be $\mathcal{T}_c$ continuous and one to one, the inverse map $F \to \theta(F)$ defined on $S = \bigcup_j S_j$ is also continuous when restricted to S_j.

Therefore, given ϵ and j there exists a symmetric $\mathcal{T}_c$ neighborhood of zero, say $V(\epsilon, j)$ such that if F and G belong to S_j and $F - G \in V(\epsilon, j)$, then $\|\theta(G) - \theta(F)\| < \epsilon$. Let $N(\epsilon, j)$ be so large that for $n \geqq N(\epsilon, j)$ one has

$$P\{F(n, Z) \in F + \tfrac{1}{8} V(\epsilon, j) \mid F\} > 1 - \epsilon \tag{33}$$

for every $F \in S$. Let $\hat{F}(n, z, \epsilon, j)$ be an arbitrary point of $S_j \cap [F(n, z) + V(\epsilon, j)/4]$ if this set is not empty, and some arbitrary point of S_j otherwise. Then

$$P^*\{\|\theta(\hat{F}) - \theta(F)\| > \epsilon \mid F\} < \epsilon \tag{34}$$

if $n \geqq N(\epsilon, j)$ and $F \in S_j$.

Let now $M_j = N(1/j, j)$, define $\nu_n = j$ if $M_j < n \leqq M_{j+1}$ and let $\delta_n = 1/\nu_n$. Define an estimate T_n by

$$T_n = \theta[\hat{F}(n, z, \delta_n, \nu_n)]. \tag{35}$$

Then for every n and every $\theta \in K_{\nu_n}$ we have

$$P^*\{\|T_n - \theta\| \geqq \delta_n \mid \theta\} \leqq \delta_n. \tag{36}$$

This completes the proof of the lemma except for the fact that T_n need not be measurable.

One can achieve measurability by giving a definite rule of selection for $\hat{F}$. Although this is not difficult it is rather long and uninteresting and will be omitted.

Instead of proceeding as above, one could try to show that maximum likelihood estimates taken over a compact K_j are uniformly consistent for points of K_j and then apply a diagonal process. This results rather easily from lemmas 1 and 2 of the preceding section, at least when assumptions (A) are satisfied. The preceding lemma is however of a more general applicability.

The sequence $\{K_j\}$ of compacts used in the proof of lemma 4 will be used again in the sequel. In this respect we shall use the following notation. Let δ_j' be a decreasing sequence of numbers such that δ_j' is smaller than the distance between K_j and K_{j+1}^c. Furthermore, choose the spheres V_θ of assumption 7 in such a way that their radii stay bounded from below when θ belongs to K_j. This is possible because of the compactness of K_j. Let δ_j'' be the lower bound of the radius of the spheres V_θ when θ belongs to K_j and let $\delta_j = \min(\delta_{j+1}', \delta_{j+1}'')$. Under nice circumstances $\{\delta_j\}$ does not necessarily tend to zero as j increases. Furthermore, for every $t \in \Theta$, let $\nu(t)$ be the smallest integer j for which t belongs to K_j.

LEMMA 5. *Let assumptions (A) be satisfied. Then, there exists a sequence $\{T_n\}$ of measurable strict estimates having the following property. For every compact $K \subset \Theta$ and every $\epsilon > 0$ there exists a number b such that*

$$\sup_n \sup_{\theta \in K} P\{\sqrt{n}\|T_n - \theta\| \geq b \mid \theta\} < \epsilon . \tag{37}$$

PROOF. Let U_j be the sphere of radius $\delta_j/4$ centered at the origin. Let $\{\xi_n\}$ be a sequence of strict estimates, uniformly consistent on the compacts of Θ. If $\nu(\xi_n) = j$, let $S(\xi_n) = \xi_n + U_j \cap \Theta$. Clearly $S(\xi_n) \subset K_{\nu(\xi_n)+1}$. Take for T_n any value $t \in S(\xi_n)$ such that

$$\Phi_n(z_n, t) \geq \sup_\tau \{\Phi_n(z_n, \tau);\ \tau \in S(\xi_n)\} - \frac{1}{n}. \tag{38}$$

If T_n is measurable, it has the desired property. To prove this, let θ be a particular point of Θ. Assume $\nu(\theta) = j$. Then for $\tau \in \theta + 4U_j$ we have

$$\Phi_n(z_n, \tau) = \Phi_n(z_n, \theta) + n\, A_n(\theta)(\tau - \theta) - \frac{n}{2}(\tau - \theta)' M_n(\theta)(\tau - \theta), \tag{39}$$

where $M_n(\theta)$ is an average of $B_n(t)$ between θ and τ. This can also be written as

$$\Phi_n(z_n, \tau) = \Phi_n(z_n, \theta) + \frac{n}{2} Y_n' M_n^{-1}(\theta) Y_n \tag{40}$$
$$- \frac{n}{2}[\tau - \theta - M_n^{-1}(\theta) Y_n]' M_n(\theta)[\tau - \theta - M_n^{-1}(\theta) Y_n],$$

with $Y_n = A_n'(\theta)$.

Now for every $\theta \in S_j$ we have

$$P_*\{\|\xi_n - \theta\| < \tfrac{1}{16}\delta_j \mid \theta\} > 1 - \alpha_n(j). \tag{41}$$

$$P_*\{\text{for every } t \in \theta + 4U_j \text{ the inequalities } B_n(t) \geq \tfrac{1}{2}I(\theta) \text{ and } \|B_n(t)\| \leq 2\|\Gamma(\theta)\| \text{ hold} \mid \theta\} > 1 - \alpha_n'(j), \tag{42}$$

with α_n and α_n' tending to zero as n tends to infinity. Neglecting cases of "small" probability we can assume that the relations in braces are satisfied. In such a case $\xi_n \in \theta + U_j/4$ and $S(\xi_n) \subset \theta + 5U_j/4$. Consequently, writing t instead of T_n to simplify, we have

$$(43)\quad -\frac{n}{2} Y_n' M_n^{-1}(t) Y_n + \frac{n}{2}[t-\theta-M_n^{-1}(t)]' M_n(t)[t-\theta-M_n^{-1}(t) Y_n]$$
$$\leqq \frac{1}{n} - \frac{n}{2}\frac{1}{4} Y_n' \Gamma^{-1}(\theta) Y_n + \inf_{\tau \in S(\xi_n)} \frac{n}{2}[\tau-\theta-M_n^{-1}(\tau) Y_n]' M_n(\tau)$$
$$\cdot [\tau-\theta-M_n^{-1}(\tau) Y_n].$$

Since $S(\xi_n)$ contains θ this implies

$$(44)\quad \frac{n}{4}[t-\theta-M_n^{-1}(t) Y_n]'\Gamma(\theta)[t-\theta-M_n^{-1}(t) Y_n]$$
$$\leqq \frac{n}{2}[t-\theta-M_n^{-1}(t) Y_n]' M_n(t)[t-\theta-M_n^{-1}(t) Y_n]$$
$$\leqq \frac{1}{n} + 2n\|\Gamma^{-1}(\theta)\| Y_n' Y_n - \frac{n}{8} Y_n' \Gamma^{-1}(\theta) Y_n + \frac{n}{2} Y_n' M_n^{-1}(\theta) Y_n.$$

According to Markov's inequality we have

$$(45)\qquad P\{n Y_n' Y_n > a \mid \theta\} \leqq \frac{\text{trace}\,\Gamma(\theta)}{a}$$

and similar bounds for the other terms in the last expression. Therefore, there exist numbers b_1 and N such that

$$(46)\qquad P^*\{\sqrt{n}\|T_n-\theta-M_n^{-1}(T_n) Y_n\| > b_1 \mid \theta\} < \epsilon$$

for every $n \geqq N$. The numbers b_1 and N can be chosen independent of θ if θ is restricted to K_j. Therefore, by Minkowski's inequality there exist numbers b_j and N_j such that $P^*\{\sqrt{n}\|T_n-\theta\| > b_j|\theta\} < \epsilon$ if $n \geqq N_j$ and $\theta \in K_j$.

If T_n is measurable one can choose b_j such that $P\{\sqrt{n}\|T_n-\theta\| > b_j'|\theta\} < \epsilon$ for every $n < N_j$ and every $\theta \in K_j$. This follows from the fact that the map $P_\theta \to \mathcal{L}(T_n|\theta)$ is then continuous in the sense of the norm so that the set $\{\mathcal{L}(T_n|\theta); \theta \in K_j\}$ is compact in the sense of the norm and further T_n takes its values in a locally compact space Θ.

To show that T_n can be chosen measurable, assume that ξ_n itself is measurable. It is then possible to give a specific rule for the choice of T_n and show that this leads to measurable estimates. For the same reasons as in lemma 4 we will omit the details of this selection.

LEMMA 6. *If the assumptions (A) are satisfied, there exists a sequence $\{T_n\}$ of measurable estimates (not necessarily strict) such that for every positive ϵ the probabilities*

$$(47)\qquad P^*\{\sqrt{n}\|T_n-\theta-\Gamma^{-1}(\theta) A_n'(\theta)\| > \epsilon \mid \theta\}$$

tend to zero uniformly on the compacts of Θ as n tends to infinity.

PROOF. According to lemma 5 there exists a sequence $\{\xi_n\}$ of measurable strict estimates such that for every positive ϵ and every compact K one can find a number b such that

$$(48)\qquad \sup_{\theta \in K} P\{\sqrt{n}\|\xi_n-\theta\| > b \mid \theta\} < \epsilon.$$

Let $T_n = \xi_n + \Gamma^{-1}(\xi_n)A_n'(\xi_n)$. Then, if ξ_n differs little from θ we have

$$(49)\quad \sqrt{n}\,(T_n - \theta) = \Gamma^{-1}(\xi_n)\,\sqrt{n}A_n'(\theta) + \sqrt{n}\,[I - \Gamma^{-1}(\xi_n)M_n]\,(\xi_n - \theta)$$

with M_n an average of $B_n(t)$ between ξ_n and θ. Therefore,

$$(50)\quad \|\sqrt{n}\,(T_n - \theta) - \Gamma^{-1}(\theta)\,\sqrt{n}A_n'(\theta)\| \\ \leqq \sqrt{n}\|A_n'(\theta)\|\,\|\Gamma^{-1}(\theta) - \Gamma^{-1}(\xi_n)\| + \|I - \Gamma^{-1}(\xi_n)M_n\|\,\sqrt{n}\,(\xi_n - \theta)\,.$$

By lemma 1, $\Gamma^{-1}(\xi_n)M_n$ tends to I in probability, uniformly on the compacts of Θ. To obtain the desired result it is sufficient to show that T_n is measurable. This however is a consequence of the fact that $A(x, t)$, being continuous in t and measurable in x, is jointly measurable in (x, t).

The preceding lemma has an important consequence concerning the limiting distribution of $\sqrt{n}\,(T_n - \theta)$. By lemma 2 the quantity $\sqrt{n}\,A_n'(\theta)$ is asymptotically normal and according to Slutsky's theorems the same is true of $\sqrt{n}\,(T_n - \theta) = \sqrt{n}\,(\xi_n - \theta) + \Gamma^{-1}(\xi_n)\sqrt{n}\,A_n'(\xi_n)$. More precisely, according to the uniform versions of Slutsky's theorem $\mathcal{L}\{\sqrt{n}\,(T_n - \theta)\,|\,\theta\}$ converges to $\mathcal{N}[0, \Gamma^{-1}(\theta)]$ in the $\mathcal{T}_c$ sense and this uniformly on the compact subsets of Θ. This implies that f being a bounded real-valued function defined on the Euclidean space $\mathcal{E}$ and having a set of discontinuities of Lebesgue measure zero, the expectations $E\{f[\sqrt{n}(T_n - \theta)]\,|\,\theta\}$ tend to $\int f(x)\,d\mathcal{N}[0, \Gamma^{-1}(\theta)]$ and this uniformly on the compact subsets of Θ.

5. Asymptotically sufficient estimates

In many respects the construction given for the estimates of the preceding section is not very appealing. In most practical cases other estimates occur in natural manner, for instance, by using the method of moments or other methods. For this reason we shall use not only the estimates defined in the preceding section, but any sequence $\{t_n\}$ of estimates belonging to the class $\mathcal{C}$ defined below.

DEFINITION. *A sequence $\{t_n\}$ of estimates will be said to belong to the class $\mathcal{C}$ if it satisfies the following conditions:*

(1) *For each n, the estimate t_n is a measurable strict estimate.*

(2) *For every positive ϵ and every compact $K \subset \Theta$ there is a number c such that*

$$(51)\qquad \sup_n \sup_{\theta \in K} P\{\sqrt{n}\|t_n - \theta\| \geqq c\,|\,\theta\} < \epsilon\,.$$

From an estimate t_n it is possible to obtain other estimates T_n and τ_n defined as follows.

Let $Y_n(\xi) = \Gamma^{-1}(\xi)A_n'(\xi)$. Define T_n by $T_n = t_n + Y_n(t_n)$, and let τ_n be a function of T_n such that $\tau_n(T_n) \in \Theta$ and

$$(52)\qquad \|T_n - \tau_n\| \leqq 2\inf_{\xi}\{\|T_n - \xi\|;\ \xi \in \Theta\}. \quad + \frac{1}{n^2}$$

Note that when t_n is measurable, the same is true of T_n and then τ_n can be chosen to be measurable.

The class of sequences $\{T_n\}$ of estimates of the form $T_n = t_n + Y_n(t_n)$ with $\{t_n\} \in \mathcal{C}$ will be denoted by $\mathfrak{D}$.

For further use note that if $\{T_n\} \in \mathfrak{D}$ and τ_n is measurable so is the quantity $\Delta_n = \tau_n + Y_n(\tau_n) - T_n$.

THEOREM 1. *Let assumptions (A) be satisfied and let $\{T_n\}$ be a sequence of estimates belonging to the class $\mathfrak{D}$. Then the sequence $\{T_n\}$ is asymptotically sufficient in the sense that there exist nonnegative functions $q_n(z_n, \theta)$, each the product of a function of z_n only by a function of T_n and θ only such that*

$$\int^* |p_n(z_n, \theta) - q_n(z_n, \theta)| \, d\nu^n \tag{53}$$

tends to zero uniformly on the compacts of Θ as n tends to infinity.

Note. The assumptions of measurability of t_n, T_n and τ_n are irrelevant to the present theorem.

PROOF. According to assumptions (A), for every θ such that $\|\theta - \tau_n\| \leqq \delta_{\nu(\tau_n)}$ one can write

$$\Phi_n(z_n, \theta) = \Phi_n(z_n, \tau_n) + nA_n(\tau_n)(\theta - \tau_n) - \frac{n}{2}(\tau_n - \theta)' M_n(\tau_n, \theta)(\tau_n - \theta), \tag{54}$$

where $M_n(\tau_n, \theta)$ denotes the usual average of $B_n(\xi)$ between τ_n and θ. Simple rearrangements show that this can also be written as

$$\Phi_n(z_n, \theta) = \Psi_n(z_n) - \frac{n}{2}(T_n - \theta)'\Gamma(\theta)(T_n - \theta) + R_n[z_n T_n, \tau_n, \theta] \tag{55}$$

with

$$\Psi_n(z_n) = \Phi_n(z_n, \tau_n) + \frac{n}{2} Y_n'(\tau_n)\Gamma(\tau_n) Y_n(\tau_n) - \frac{n}{2}\Delta_n'\Gamma(\tau_n)\Delta_n \tag{56}$$

and

$$\begin{aligned} R_n[z_n, T_n, \tau_n, \theta] = & -\frac{n}{2}(\tau_n - \theta)'[M_n(\tau_n, \theta) - \Gamma(\tau_n)](\tau_n - \theta) \\ & - \frac{n}{2}(T_n - \theta)'[\Gamma(\tau_n) - \Gamma(\theta)](T_n - \theta) - n\Delta_n'\Gamma(\tau_n)(T_n - \theta). \end{aligned} \tag{57}$$

Our purpose is to show that this residual is eventually small and then remove it from further considerations. The property assumed for $\{t_n\}$ implies that for each integer j there is a sequence $\{c'_{n,j}\}$ of real numbers such that $4c'_{n,j} \leqq \delta_{j+1} n^{1/4}$ and such that

$$\sup_{\theta \in K_{j+1}} P^*\{\sqrt{n}\|t_n - \theta\| \geqq c'_{n,j} | \theta\} \tag{58}$$

tends to zero as n tends to infinity.

Let

$$\begin{cases} D_n' = \sup_\xi \left\{ \|I - \Gamma^{-1}(t_n) B_n(\xi)\|;\ \|\xi - t_n\| \leqq \frac{c'_{n,\nu(t_n)}}{\sqrt{n}} \right\} \\ D_n'' = \sup_\xi \left\{ \|I - \Gamma^{-1}(t_n)\Gamma(\xi)\|;\ \|\xi - t_n\| \leqq \frac{c'_{n,\nu(t_n)}}{\sqrt{n}} \right\} \\ D_n = \max[D_n', D_n'']. \end{cases} \tag{59}$$

According to lemma 1, for each j there exists a sequence $\{a_{n,j}\}$ tending to zero and such that $\sup[P^*\{D_n \geqq a_{n,j} | \theta\};\ \theta \in K_{j+1}]$ tends to zero as n tends to infinity. Let $c_{n,j} = \min\{c'_{n,j}, a_{n,j}^{-1/2}\}$. Clearly

$$\sup_{\theta \in K_{j+1}} P^*\{\sqrt{n}\|t_n - \theta\| \geqq c_{n,j} | \theta\} \tag{60}$$

also tends to zero as n tends to infinity. From this we can deduce that t_n, T_n and τ_n are reasonably close to each other. By definition $T_n - t_n = \Gamma^{-1}(t_n)A'_n(t_n)$. Let j be an arbitrary integer and let θ be a point of K_j considered as "true value." If $\sqrt{n}\|t_n - \theta\| \leqq c_{n,j}$, Taylor's formula can be applied to $A'_n(t_n)$ giving

$$(61) \qquad A'_n(t_n) = A'_n(\theta) - \tilde{M}_n(t_n, \theta)(t_n - \theta)$$

with $\tilde{M}_n(t_n, \theta)$ an average of $B_n(\xi)$ between t_n and θ. Therefore,

$$(62) \qquad \|T_n - t_n\| \leqq \|\Gamma^{-1}(t_n)\Gamma(\theta)\Gamma^{-1}(\theta)A'_n(\theta)\| + \|[\Gamma^{-1}(t_n)\tilde{M}_n(t_n, \theta) - I](t_n - \theta)\| + \|t_n - \theta\|$$

$$(63) \qquad \|T_n - t_n\| \leqq D_n\|\Gamma(\theta)A'_n(\theta)\| + (D_n + 1)\|t_n - \theta\| .$$

Let β_j be a bound for $\|\Gamma(\theta)\|^{1/2}$ over K_{j+1}. According to Chebyshev's inequality

$$(64) \qquad P\{\sqrt{n}\|\Gamma^{1/2}(\theta)A'_n(\theta)\| \geqq a_{n,j} | \theta\} \leqq \frac{r}{a^2_{n,j}}.$$

Consequently, if $\{a_{n,j}\}$ is any sequence of numbers tending to infinity we have

$$(65) \qquad \sqrt{n}\|T_n - t_n\| \leqq \beta_j a_{n,j}\, a_{n,j} + (a_{n,j} + 1)\, c_{n,j}$$

except for cases of probability tending to zero uniformly on K_j. Take for $a_{n,j}$ the value $a_{n,j} = \min [a_{n,j}^{-1/2}, (1/4)\delta_{j+1}n^{1/2}]$ and let $\gamma_{n,j}$ be the quantity obtained by substituting this value in

$$(66) \qquad \min\{[\beta_j\, a_{n,j}\, a_{n,j} + (a_{n,j} + 1)\, c_{n,j}],\ \tfrac{1}{4}\delta_{j+1}\sqrt{n}\}.$$

Note that $\gamma_{n,j}n^{-1/2}$ tends to zero as n tends to infinity. The preceding inequalities imply that the following three quantities

$$(67) \qquad \sup [P^*\{\sqrt{n}\|T_n - t_n\| \geqq \gamma_{n,j} | \theta\};\ \theta \in K_j]$$

$$(68) \qquad \sup [P^*\{\sqrt{n}\|T_n - \tau_n\| \geqq 2\gamma_{n,j} | \theta\};\ \theta \in K_j]$$

$$(69) \qquad \sup [P^*\{\sqrt{n}\|t_n - \tau_n\| \geqq 3\gamma_{n,j} | \theta\};\ \theta \in K_j]$$

tend to zero as n tends to infinity.

A similar computation can be applied to the quantity Δ_n. Using a Taylor expansion around τ_n one obtains

$$(70) \qquad -\Delta_n = [I - \Gamma^{-1}(t_n)\tilde{M}(\tau_n, t_n)](t_n - \tau_n) + [\Gamma^{-1}(t_n) - \Gamma^{-1}(\tau_n)]\, A'_n(\tau_n).$$

It is clear from this formula that $\sqrt{n}\Delta_n$ tends to zero in probability as n tends to infinity. By arguments similar to the ones used above one can show that for each j there exists a sequence $\{\zeta_{n,j}\}$ tending to zero as n increases and such that

$$(71) \qquad \sup [P^*\{\sqrt{n}\|\Gamma^{1/2}(\tau_n)\Delta_n\| \geqq \zeta_{n,j} | \theta\};\ \theta \in K_{j+1}]$$

tends to zero as n tends to infinity.

For each integer j let $\{b'_{n,j}\}$ be a sequence such that $b'_{n,j}\zeta_{n,j} \leqq 1$ and such that $b'_{n,j}$ tends to infinity as n increases. Let $\zeta_n(\tau) = \zeta_{n,\nu(\tau)}$ and $b'_n(\tau) = b'_{n,\nu(\tau)}$. Assume also that $\|\Gamma^{1/2}(\tau)\|b'_n(\tau) \leqq (1/4)\delta_{\nu(\tau)}\sqrt{n}$ and that $b'_n(\tau)n^{-1/2}$ tends to zero as n tends to infinity.

Let

$$(72)\quad \begin{cases} \bar{D}'_n(z_n) = \sup_{\xi} \{ \|\Gamma^{-1}(\tau_n) B_n(\xi) - I\|;\ \|\xi - \tau_n\| \leq \|\Gamma^{1/2}(\tau_n)\|\, b'_n(\tau_n)\, n^{-1/2}\} \\ \bar{D}''_n(z_n) = \sup_{\xi} \{ \|\Gamma^{-1}(\tau_n) \Gamma(t) - I\|;\ \|\xi - \tau_n\| \leq \|\Gamma^{1/2}(\tau_n)\|\, b'_n(\tau_n)\, n^{-1/2}\} \\ \bar{D}_n = \max\{ \bar{D}'_n(z_n),\ \bar{D}''_n(z_n)\}. \end{cases}$$

For each integer j, there exists a sequence $\{\bar{a}_{n,j}\}$ tending to zero and such that

$$(73)\qquad \sup\,[P^*\{\bar{D}_n \geq \bar{a}_{n,j} \mid \theta\};\ \theta \in K_{j+1}]$$

tends to zero as n tends to infinity. Let

$$(74)\qquad b_n(\tau) = \min\{ b'_n(\tau),\ [\bar{a}_{n,\nu(\tau)}]^{-1/4}\}.$$

The conditions

$$(75)\qquad \begin{aligned} \sqrt{n}\|\Gamma^{1/2}(\tau_n)(T_n - \theta)\| &\leq b_n(\tau_n) \\ \sqrt{n}\|\Gamma^{1/2}(\tau_n)(\tau_n - \theta)\| &\leq b_n(\tau_n) \end{aligned}$$

imply

$$(76)\qquad |R_n(z_n, T_n, \tau_n, \theta)| \leq 2\bar{D}_n b_n^2(\tau_n) + b_n(\tau_n)\|\Gamma^{1/2}(\tau_n)\Delta_n\sqrt{n}\|.$$

Finally, for every positive ϵ and every integer j the quantity

$$(77)\qquad \sup_{\theta \in K_j} P^*\{2\bar{D}_n b_n^2(\tau_n) + b_n(\tau_n)\|\Gamma^{1/2}(\tau_n)\Delta_n\sqrt{n}\| \geq \epsilon \mid \theta\}$$

tends to zero as n tends to infinity. Using the diagonal process, we can find a sequence $\{\epsilon_n\}$ tending to zero and such that

$$(78)\qquad \sup_{\theta \in K_j} P^*\{2\bar{D}_n b_n^2(\tau_n) + b_n(\tau_n)\|\Gamma^{1/2}(\tau_n)\Delta_n\sqrt{n}\| \geq \epsilon_n \mid \theta\}$$

tends to zero for every integer j.

Let χ_n be the indicator of the set S_n defined by

$$(79)\qquad S_n = \{ z_n\colon 2\bar{D}_n b_n^2(\tau_n) + b_n(\tau_n)\|\Gamma^{1/2}(\tau_n)\Delta_n\sqrt{n}\| \leq \epsilon_n\}$$

and let $v_n(T_n, \theta)$ be equal to unity if both $\sqrt{n}\|\Gamma^{1/2}(\tau_n)(T_n - \theta)\| \leq b_n(\tau_n)$ and $\sqrt{n}\|\Gamma^{1/2}(\tau_n)(\tau_n - \theta)\| \leq b_n(\tau_n)$ and let $v_n(T_n, \theta)$ be equal to zero otherwise. Let

$$(80)\quad q_n(z_n, \theta) = \chi_n(z_n)\, v_n(T_n, \theta) \exp\{\Psi_n(z_n) - \frac{n}{2}(T_n - \theta)'\,\Gamma(\theta)(T_n - \theta)\}.$$

Let $\Omega_n(\theta)$ be the subset of $\mathcal{X}_n$ where $\chi_n(z_n)v_n(T_n, \theta) = 1$. Then

$$(81)\quad \begin{aligned} \int |q_n(z_n, \theta) - p_n(z_n, \theta)|\, d\nu^n &\leq P^*\{\Omega_n^c(\theta) \mid \theta\} + \int_{\Omega_n(\theta)} |e^{\epsilon_n} - 1|\, p_n(z_n, \theta)\, d\nu^n \\ &\leq P^*\{\Omega_n^c(\theta) \mid \theta\} + e^{\epsilon_n} - 1. \end{aligned}$$

This quantity tends to zero uniformly on the compacts of Θ as n tends to infinity.

By construction, the function $q_n(z_n, \theta)$ is the product of two functions, the first depending on z_n only and the other depending on T_n and θ only. Therefore, the estimate T_n is a sufficient statistic for the family $\{q_n(z_n, \theta), \theta \in \Theta\}$. Of course, $q_n(z_n, \theta)$ is not necessarily a probability density since its norm might differ from unity. However, the difference tends to zero uniformly on the compacts of Θ, so that if necessary q_n could be normalized.

For further use, note that in $\Omega_n(\theta)$ we have taken $\log q_n(z_n, \theta) = \Psi_n(z_n) - (n/2)(T_n - \theta)'\Gamma(\theta)(T_n - \theta)$. For problems involving *a posteriori* probabilities it is often more convenient to take a q_n' defined by

(82) $$\log q_n'(z_n, \theta) = \Psi_n(z_n) - \frac{n}{2}(T_n - \theta)'\Gamma(\tau_n)(T_n - \theta)$$

for values of z_n in $\Omega_n(\theta)$. The preceding argument shows also that, as n tends to infinity,

(83) $$\int^* |p_n(z_n, \theta) - q_n'(z_n, \theta)|\, d\nu^n$$

tends to zero uniformly on the compacts of Θ.

6. Asymptotic normality

It has been shown in section 4 that the estimates $\{T_n\}$ of the class $\mathfrak{D}$ are asymptotically normally distributed. Such a result is, however, much too weak to be of general use in finding asymptotically optimal test or decision functions. It is possible to get more precise results by using the results of section 5 to form approximations to the distribution of T_n. It will be convenient to use instead of T_n the random vector $X_n = \sqrt{n}\, T_n$. To achieve an equivalent modification of the parameter space let $\bar{\theta}_n$ denote the value $\bar{\theta}_n = \theta\sqrt{n}$. The associated estimate τ_n will be considered as a function of $x \in \mathcal{E}$. Instead of using the function q_n of section 5 in the form given there it will be convenient to multiply it by $[\det \Gamma(\theta)]^{1/2} [\det \Gamma(\tau_n)]^{1/2}$. This modification is of no importance asymptotically.

Let $w_n(x, \theta)$ be equal to unity if $\|\Gamma^{1/2}(\tau_n)(x - \bar{\theta}_n)\| \leq b_n(\tau_n)$ and $\|\Gamma^{1/2}(\tau_n)\sqrt{n}[\tau_n(x) - \theta]\| \leq b_n(\tau_n)$ and let w_n be equal to zero otherwise.

Let $g_n(x, \theta)$ be the normal density

(84) $$g_n(x, \theta) = \frac{[\det \Gamma(\theta)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(x - \bar{\theta}_n)'\Gamma(\theta)(x - \bar{\theta}_n)\}.$$

Let $P_{\theta,n}$ be the actual distribution of X_n given θ calculated using the original product measures $P_\theta^{(n)}$. The approximate densities q_n and q_n' defined in section 5 give approximations to $P_{\theta,n}$ to be denoted by $Q_{\theta,n}$ and $Q_{\theta,n}'$. The form of q_n implies that

(85) $$Q_{\theta\ n}(S) = \int_S g_n(x, \theta) w_n(x, \theta)\, d\lambda_n$$

where λ_n is some measure on the Euclidean space $\mathcal{E}$ in which X_n takes its values. It is easily seen that λ_n is finite on the compacts of $\mathcal{E}$.

The approximation q_n' gives a measure $Q_{\theta,n}'$ defined by

(86) $$Q_{\theta,n}'(S) = \int g_n'(x, \theta) w_n(x, \theta)\, d\mu_n'$$

with

(87) $$g_n'(x, \theta) = \frac{[\det \Gamma(\tau_n)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(x - \bar{\theta}_n)'\Gamma(\tau_n)(x - \bar{\theta}_n)\}$$

and μ_n' some measure on the space $\mathcal{E}$.

Let λ be the Lebesgue measure on $\mathcal{E}$.

THEOREM 2. *Let assumptions (A) be satisfied. Let $\{T_n\}$ be a sequence of estimates of the class $\mathfrak{D}$. Then, the notation being as above, there exists a sequence $\{\varphi_n\}$ of measurable functions from $\mathcal{E}$ to $\mathcal{E}$ such that*

(1) $\sup_x \|\varphi_n(x) - x\|$ *tends to zero as n tends to infinity.*

(2) *Let μ_n be the measure defined by $\mu_n(S) = \lambda[\varphi_n^{-1}(S)]$, and let $\bar{Q}_{\theta,n}$ be the measure defined by*

$$\bar{Q}_{\theta,n}(S) = \int_S g_n(x, \theta)\, d\mu_n \tag{88}$$

then for every compact $K \subset \Theta$ the quantity

$$\sup_{\theta \in K} \|P_{\theta,n} - \bar{Q}_{\theta,n}\| \tag{89}$$

tends to zero as n tends to infinity.

PROOF. It has been shown in section 5 that $\|P_{\theta,n} - Q_{\theta,n}\|$ tends to zero uniformly on the compacts of Θ when n tends to infinity. Also the actual distribution $\mathcal{L}\{X_n - \theta_n|\theta\}$ of $X_n - \theta_n$ for the measure $P_{\theta,n}$ tends to the measure $\mathcal{N}[0, \Gamma^{-1}(\theta)]$ uniformly in the $\mathcal{T}_c$ sense on the compacts of Θ. This implies that, in a certain sense, λ_n tends to the Lebesgue measure λ on $\mathcal{E}$. More precisely, let $\lambda_{\theta,n}$ be the measure defined by $\lambda_{\theta,n}(B) = \lambda_n(\bar{\theta}_n + B)$. For every bounded continuous function γ defined on $\mathcal{E}$, the difference

$$\int \gamma(x - \bar{\theta}_n)\, dQ_{\theta,n} - \int \gamma(x - \bar{\theta}_n)\, dP_{\theta,n} \tag{90}$$

tends to zero uniformly on the compacts of Θ. This and the structure of the function $w_n(x, \theta)$ entering in the definition of $Q_{\theta,n}$ implies that $\lambda_{\theta,n}$ converges to the Lebesgue measure λ uniformly in the vague sense on the compacts of Θ.

For every integer m, there exists an integer n'_m such that $n \geq n'_m$ implies $\|P_{\theta,n} - Q_{\theta,n}\| < 1/2m$ for every $\theta \in K_{m+1}$.

Let a_m be a number so large that if U is normal with mean zero and covariance $\Gamma^{-1}(\theta)$ then

$$\sup [P\{\|X\| \geq a_m | \theta\};\ \theta \in K_{m+1}] \leq \frac{1}{2m}. \tag{91}$$

Let $n''_m \geq n'_m$ be so large that $n \geq n''_m$ implies

$$\sup\{Q_{\theta,n}[\|X_n - \bar{\theta}_n\| \geq a_m];\ \theta \in K_{m+1}\} < \frac{1}{m}. \tag{92}$$

Furthermore, assume $\sqrt{n''_m} \geq 4a_m/\delta_m$.

Let $K_j^{(n)}$ be the image $\sqrt{n}\, K_j$ of the set K_j in $\mathcal{E}$. Assume without loss of generality that the origin is a point of K_1. Finally let β_m be a bound for $1 + 2(2\pi)^{-r/2}[\det \Gamma(\theta)]^{1/2}$ for $\theta \in K_{m+1}$.

Let $\{\eta_m\}$ be a sequence of positive numbers decreasing to zero as m tends to infinity. Starting from the origin, pave the space $\mathcal{E}$ with cubes having sides equal to η_m. Let $W_m^{(n)}$ be the part of $\mathcal{E}$ covered by cubes whose centers are at distance less than a_m from $K_m^{(n)}$. Note that θ being any point of Θ the number of cubes at distance less than a_m from $\bar{\theta}_n$ is a bounded function of n.

Since the measures $\lambda_{\theta,n}$ converge to λ uniformly in the vague sense for $\theta \in K_{m+1}$, there exists a number $n_m \geq n''_m$ such that

$$\lambda_n(C)\left[1 - \frac{1}{m\beta_m}\right] \leq \lambda(C) \leq \left[1 + \frac{1}{m\beta_m}\right]\lambda_n(C) \tag{93}$$

for any $n \geqq n_m$ and every cube C element of the pavement giving $W_m^{(n)}$. For every such cube $\lambda_n(C)$ is positive and there exists a measurable function $\varphi_{n,m,C}$ from C to C such that

$$(94)\qquad \lambda\,[\varphi_{n,m,C}^{-1}(B)] = \frac{\lambda(C)}{\lambda_n(C)}\lambda_n(B)$$

for every Borel set $B \subset C$. For every $x \in W_m^{(n)}$ let $\varphi_{n,m}$ be defined by juxtaposition of the $\varphi_{n,m,C}$ and for x not in $W_m^{(n)}$ define $\varphi_{n,m}$ to be the identity map. Let $\mu_{n,m}(B) = \lambda[\varphi_{n,m}^{-1}(B)]$. For every Borel set $B \subset W_m^{(n)}$ we have

$$(95)\qquad \mu_{n,m}(B) = \sum_C \lambda\,[\varphi_{n,m}^{-1}(B \cap C)]$$

the summation being over the cubes $C \subset W_m^{(n)}$. From this it follows that

$$(96)\qquad \mu_{n,m}(B) = \sum_C \frac{\lambda(C)}{\lambda_n(C)}\lambda_n(B \cap C)$$

and finally

$$(97)\qquad \left[1 - \frac{1}{m\beta_m}\right]\lambda_n(B) \leqq \mu_{n,m}(B) \leqq \left[1 + \frac{1}{m\beta_m}\right]\lambda_n(B).$$

Define a measure $\bar{Q}_{\theta,n,m}$ by

$$(98)\qquad \bar{Q}_{\theta,n,m}(S) = \int_S g_n(x,\theta)\,d\mu_{n,m}.$$

It follows from the above inequalities that

$$(99)\qquad |\bar{Q}_{\theta,n,m}(S) - Q_{\theta,n}(S)| \leqq \frac{3}{m}$$

for every $\theta \in K_m$, every Borel set S and every $n \geqq n_m$. Let $\{\varphi_n\}$ be the sequence of functions defined by $\varphi_n = \varphi_{n,m}$ if $n_m < n \leqq n_{m+1}$ and similarly for the measures μ_n and $\bar{Q}_{\theta,n}$. Then

$$(100)\qquad \|\bar{Q}_{\theta,n} - Q_{\theta,n}\| \leqq \frac{6}{m};$$

hence

$$(101)\qquad \|\bar{Q}_{\theta,n} - P_{\theta,n}\| \leqq \frac{7}{m}$$

provided $n_m < n \leqq n_{m+1}$ and $\theta \in K_m$. This and the fact that $\{\eta_m\}$ tends to zero completes the proof of the theorem.

The construction of $\{\varphi_n\}$ given in the above proof has another consequence of some interest. Let $g(x, \xi, \theta)$ be the density with respect to λ of the normal distribution $\mathcal{N}[\xi, \Gamma^{-1}(\theta)]$, that is,

$$(102)\qquad g(x,\xi,\theta) = \frac{[\det \Gamma(\theta)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(x-\xi)'\Gamma(\theta)(x-\xi)\}.$$

Let φ be a function from $\mathcal{E}$ to $\mathcal{E}$ and let $\tilde{g}(x, \xi, \theta)$ be defined by $g[\varphi(x), \xi, \theta] = \tilde{g}(x, \xi, \theta)$. According to lemma 6 for every integer m there exists a positive number η'_m such that $\|\varphi(x) - x\| \leqq r\eta'_m$ implies

$$(103)\qquad \int^* |\tilde{g}(x,\xi,\theta) - g(x,\xi,\theta)|\,d\lambda < \frac{1}{m}$$

for every $\theta \in K_{m+1}$ and every ξ.

Suppose then that the sequence $\{\eta_m\}$ used in the construction of $\{\varphi_n\}$ is such that $\eta_m \leqq \eta'_m$.

Let $G_{\theta, n}$ be the normal measure

$$G_{\theta, n}(S) = \int_S g_n(x, \theta)\, d\lambda . \tag{104}$$

Assuming that n verifies the inequality $n_m < n \leqq n_{m+1}$ let

$$\tilde{G}_{\theta, n}(S) = \int_S \frac{[\det \Gamma(\theta)]^{1/2}}{(2\pi)^{r/2}} \exp \{ -\tfrac{1}{2} [\varphi_n(x) - \bar{\theta}_n]' \Gamma(\theta) [\varphi_n(x) - \bar{\theta}_n] \} d\lambda . \tag{105}$$

For every $\theta \in K_m$ one has also $\|\tilde{G}_{\theta, n} - G_{\theta, n}\| < 1/m$. Owing to the definition of $\bar{Q}_{\theta, n}$ the measure $\tilde{G}_{\theta, n}$ is related to $\bar{Q}_{\theta, n}$ by the equality

$$\bar{Q}_{\theta, n}(S) = \tilde{G}_{\theta, n}[\varphi_n^{-1}(S)] . \tag{106}$$

Therefore, if $F_{\theta, n}$ is defined by

$$F_{\theta, n}(S) = G_{\theta, n}[\varphi_n^{-1}(S)] \tag{107}$$

we have also

$$\|F_{\theta, n} - \bar{Q}_{\theta, n}\| < \frac{1}{m} . \tag{108}$$

From this one concludes

$$\|Q_{\theta, n} - F_{\theta, n}\| \leqq \frac{7}{m} \tag{109}$$

and

$$\|P_{\theta, n} - F_{\theta, n}\| \leqq \frac{8}{m} \tag{110}$$

for every $\theta \in K_m$ and n such that $n_m < n \leqq n_{m+1}$. This gives the following corollary.

COROLLARY. *Let X_n be $\sqrt{n}\, T_n$ as before and let U_n have a normal distribution with expectation $\bar{\theta}_n = \sqrt{n}\theta$ and covariance matrix $\Gamma(\theta)$. Let $\{\varphi_n\}$ be as above. Then for every compact $K \subset \Theta$ the quantity*

$$\sup_{\theta \in K} \|\mathcal{L}(X_n \mid \theta) - \mathcal{L}\{\varphi_n(U_n) \mid \theta\}\| \tag{111}$$

tends to zero as n tends to infinity.

This corollary is weaker than the corresponding result of Wald [3], lemma 2, in the sense that Wald obtains uniform convergence on the whole of Θ where we obtain only uniform convergence on the compacts of Θ. Nothing more can be expected under assumptions (A). However, this corollary is stronger than Wald's lemma 2 in the sense that where Wald obtains only a set transformation we have obtained a point transformation. The difference is of great importance when problems of estimation are considered.

7. Some consequences of a general nature

The general implications of the results of the preceding sections can best be described in the framework of decision theory. To be able to apply theorems 1 and 2 directly, we shall restrict our attention to situations where for each n the space D_n of available decisions is a Borel subset of a Euclidean space. Furthermore, it will be assumed throughout that, for each n, the loss function W_n defined on $D_n \times \Theta$ is jointly measurable in its arguments and that there exists a finite number M such that $|W_n(d, \theta)| \leqq M$ for every triplet (n, d, θ). Only measurable decision functions will be considered, but no restriction will be placed on the extent to which decision functions are randomized. To be precise, a decision function depending on n observations is a function $z_n \to F_{z_n}$ from the product space $\mathcal{X}_n$ to the space of probability measures on the Borel subsets of D_n. Such a decision function is called measurable if for every closed set S of D_n the function $z_n \to F_{z_n}(S)$ is a measurable function of z_n.

Let f_n be a decision function depending on z_n. It will be convenient to denote by $\bar{W}_n(f_n, \theta)$ the integral

$$\bar{W}_n(f_n, \theta) = \int_{D_n} W_n(\xi, \theta)\, df_n(\xi) \tag{112}$$

and by $R_n(f_n, \theta)$ the corresponding risk function

$$R_n(f_n, \theta) = E\{\bar{W}_n(f_n, \theta) \mid \theta\}. \tag{113}$$

Let $\{T_n\}$ be a sequence of estimates of the class $\mathfrak{D}$. According to theorem 1, for every sequence $\{f_n\}$ of decision functions depending on z_n, there is a sequence $\{\bar{f}_n\}$ of decision functions depending on $\{T_n\}$ only such that $R_n(\bar{f}_n, \theta) - R_n(f_n, \theta)$ tends to zero as n tends to infinity, and this uniformly on the compacts of Θ.

More precisely, let M be a bound for the sequence $\{W_n\}$ and let $a_n(\theta)$ be the integral

$$a_n(\theta) = \int | p_n(z_n, \theta) - q_n(z_n, \theta) |\, d\nu^n. \tag{114}$$

To every f_n one can associate $\bar{f}_n$ depending on T_n only in such a way that

$$|R_n(\bar{f}_n, \theta) - R_n(f_n, \theta)| \leqq M a_n(\theta), \tag{115}$$

for every $\theta \in \Theta$.

Let $\mathcal{M}$ be the space of bounded signed measures on Θ. The above inequality implies that

$$\sup_{f_n} \left| \int R_n(f_n, \theta)\, d\sigma - \int R_n(\bar{f}_n, \theta)\, d\sigma \right| \tag{116}$$

tends to zero as n tends to infinity, uniformly for σ in the $\mathcal{T}_c$ compact subsets of $\mathcal{M}$.

When considering decision procedures $\bar{f}_n$ which depend on $\{T_n\}$ only, it is possible to define several risk functions. The exact distribution $P_{\theta, n}$ of $X_n = \sqrt{n} T_n$ leads to the risk denoted by $R_n(\bar{f}_n, \theta)$. The approximations $\bar{Q}_{\theta, n}$ and $Q'_{\theta, n}$ lead to risks to be denoted by $\bar{R}_n(\bar{f}_n, \theta)$ and $R'_n(\bar{f}_n, \theta)$, respectively. Clearly, the quantity

$$\sup_{\bar{f}_n} |\bar{R}_n(\bar{f}_n, \theta) - R_n(\bar{f}_n, \theta)| \tag{117}$$

tends to zero as n tends to infinity, uniformly on the compacts of Θ, and similarly for R'_n.

Let $\{\sigma_n\}$ be a sequence of *a priori* distributions on Θ. By definition a sequence $\{f_n\}$ of decision procedures is asymptotically Bayes with respect to $\{\sigma_n\}$ if the difference

$$\int R_n(f_n, \theta)\, d\sigma_n - \inf_{g_n} \int R_n(g_n, \theta)\, d\sigma_n \tag{118}$$

tends to zero as n tends to infinity.

In this definition f_n and g_n denote measurable decision procedures depending on the sample point z_n.

To simplify the statement of the next proposition, let us say that a set S of bounded signed measures on Θ is "tight" if the norms $\|\sigma\|$, $\sigma \in S$, form a bounded set and if for every $\epsilon > 0$ there is a compact $K \subset \Theta$ such that $|\sigma|(K^c) < \epsilon$ for every $\sigma \in S$. It is equivalent to say that S is $\mathcal{T}_c$ relatively compact in $\mathcal{M}$.

Assuming, as always, that $\{W_n\}$ is a bounded measurable sequence one obtains the following result.

PROPOSITION 1. *Let assumptions (A) be satisfied. Let $\{\bar{T}_n\}$ be a sequence of estimates of the class $\mathfrak{D}$. Let $\{\sigma_n\}$ be a tight sequence of measures on Θ and let $\{\bar{f}_n\}$ be a sequence of decision functions depending on $\{\bar{T}_n\}$ only.*

A necessary and sufficient condition that $\{\bar{f}_n\}$ be asymptotically Bayes with respect to $\{\sigma_n\}$, when the distributions considered are the original product measures $P_\theta^{(n)}$ on $\mathfrak{X}_n$, is that $\{\bar{f}_n\}$ be asymptotically Bayes with respect to $\{\sigma_n\}$ when the distribution of $\sqrt{n}\,\bar{T}_n$ is given by $\bar{Q}_{\theta, n}$. The statement is also valid if $\bar{Q}_{\theta, n}$ is replaced by $\bar{Q}'_{\theta, n}$.

Let σ_n be an arbitrary probability measure on Θ and let $g_n(x, \theta)$ be the density

$$(119)\quad g_n(x, \theta) = \frac{[\det \Gamma(\theta)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(x - \theta\sqrt{n})'\Gamma(\theta)(x - \theta\sqrt{n})\}.$$

It follows from Fubini's theorem that, except perhaps for a set of values of x of $\lambda + \mu_n$ measure zero, the integral

$$(120)\qquad J_n(x) = \int_\Theta g_n(x, \theta)\, d\sigma_n$$

is finite and positive. For every x for which $J_n(x) < \infty$ define a measure $\bar{\sigma}_{n, x}$ by

$$(121)\qquad \bar{\sigma}_{n, x}(B) = [J_n(x)]^{-1} \int_B g_n(x, \theta)\, d\sigma_n.$$

These measures can be used as *a posteriori* distributions of θ given $X_n = x$ for the measures $\bar{Q}_{\theta, n}$. Obviously, $\bar{\sigma}_{n, x}$ can also be considered as an *a posteriori* distribution of θ given $X_n = x$ for the normal distributions $G_{\theta, n} = \mathcal{N}\{\theta\sqrt{n}, \Gamma^{-1}(\theta)\}$. This implies that the two families $\{\bar{Q}_{\theta, n}\}$ and $\{G_{\theta, n}\}$ have in common a large class of asymptotic Bayes solutions. Let us say that a sequence $\{\bar{f}_n\}$ of measurable decision functions is regularly asymptotically Bayes for the particular choice $\{\bar{\sigma}_{n, x}\}$ if for some sequence $\{\epsilon_n\}$ tending to zero

$$(122)\qquad \int \bar{W}(\bar{f}_n, \theta)\, d\bar{\sigma}_{n, x} - \inf_{\bar{f}_n} \int \bar{W}(\bar{f}'_n, \theta)\, d\bar{\sigma}_{n, x} \leqq \epsilon_n.$$

Obviously regularly asymptotically Bayes sequences are asymptotically Bayes but not conversely. The above argument shows that regularly asymptotically Bayes sequences are asymptotically Bayes for the two families $\{\bar{Q}_{\theta, n}\}$ and $\{G_{\theta, n}\}$ simultaneously.

The *a posteriori* distribution $\bar{\sigma}_{n, x}$ are sometimes difficult to use, because along certain sequences $\{\theta_m\}$ the eigen values of the matrices $\Gamma(\theta_m)$ might become unreasonably different. To avoid this, it is often simpler to use the measures $Q'_{\theta, n}$ and the *a posteriori* distributions $\sigma'_{n, x}$ defined by

$$(123)\quad \sigma'_{n, x}(B) = \left\{\int_\Theta g'_n(x, \theta) w_n(x, \theta)\, d\sigma_n\right\}^{-1} \int_B g'_n(x, \theta) w_n(x, \theta)\, d\sigma_n.$$

For these distributions no such difficulty can occur.

The interest of asymptotic Bayes solutions lies in the fact that they possess remarkable optimal asymptotic properties. Let assumptions (A) be satisfied and let $\{W_n\}$ be a bounded sequence of measurable loss functions. Let $\{\sigma_n\}$ be a sequence of *a priori* distributions tending to a distribution σ in the sense that $\|\sigma_n - \sigma\|$ tends to zero. It can be shown that a sequence $\{f_n\}$ is asymptotically Bayes with respect to $\{\sigma_n\}$ if and only if it is asymptotically Bayes with respect to every sequence $\{\rho_n\}$ tending for the norm to a positive measure ρ absolutely continuous with respect to σ.

This implies as a particular case the following: Let $\{f_n\}$ be asymptotically Bayes with respect to the sequence $\{\sigma_n\}$ tending for the norm to σ. Let $\{f'_n\}$ be an arbitrary sequence of decision functions. If the difference $R_n(f'_n, \theta) - R_n(f_n, \theta)$ converges pointwise to a limit $v(\theta)$ then $v(\theta)$ is nonnegative except maybe on a set of values of θ of σ measure zero.

The above propositions have been proved in [9] under conditions considerably weaker than assumptions (A).

For these general conditions to be satisfied, it is sufficient but by no means necessary that assumptions (1), (2), (3) of (A) hold and that for each Borel set $B \subset \mathcal{X}$ the map $\theta \rightarrow P_\theta(B)$ be Borel measurable in θ.

It might seem plausible that under the stronger assumptions (A), results much stronger than the above propositions would be available. Unfortunately, this does not seem to be the case in general as can be judged from the results available if the family $\{P_\theta, \theta \in \Theta\}$ is the family of normal distributions with covariance matrix the identity and expectation θ a point of, say, the unit sphere in r dimensions.

By combining the results of [9] with those of the present paper one sees easily that the Bernstein-von Mises phenomenon still takes place to a certain extent. Let Θ be an open set and let σ be a measure having a positive continuous density with respect to the Lebesgue measure. Let σ_{n, z_n} be the *a posteriori* distribution of $\sqrt{n}(\theta - T_n)$ given z_n and let $G_n(\tau_n)$ be a normal distribution $G_n(\tau_n) = \mathcal{N}\{0, \Gamma^{-1}(\tau_n)\}$. Then for every $\epsilon > 0$ the quantity

$$P\{\|\sigma_{n, z_n} - G_n(\tau_n)\| > \epsilon \mid \theta\} \tag{124}$$

tends to zero as n tends to infinity and this for every $\theta \in \Theta$ except maybe for a set of values of θ of Lebesgue measure zero. Therefore, the results of [10] apply, *mutatis mutandis*, to the present situation. The same results can naturally be obtained by considering the *a posteriori* distributions $\bar{\sigma}_{n, z}$ defined by means of the measures $\bar{Q}_{\theta, n}$.

In the introduction it has been mentioned that when an estimate $\{T_n\}$ of the class $\mathfrak{D}$ is known, the methods used by Neyman in the study of B.A.N. estimates yield a variety of other estimates or tests with asymptotically desirable properties. Before we pass to an application of a much more specific character let us mention examples of results obtainable in this manner.

Suppose that at some stage in building a model of a natural phenomenon, it has been found adequate to assume that the class of probability measures under consideration is $\{P_\theta, \theta \in \Theta\}$ satisfying assumptions (A). It might happen that, at some further stage, one believes that the class corresponding to a subset Θ_1 of Θ is adequate. One might then want to test the hypothesis that θ belongs to Θ_1 against the alternative that θ belongs to $\Theta \cap \Theta_1^c$. Even more, one might desire to test the hypothesis that θ belongs to a subset Θ_2 of Θ_1 against the alternative $\theta \in \Theta_1 \cap \Theta_2^c$. Also, one might wish to obtain estimates of θ taking their values in Θ_1 or confidence intervals for θ.

If an estimate $\{T_n\}$ of the class $\mathfrak{D}$ is available, such problems can be attacked by the methods of [1]. Of course, for the B.A.N. methods to be useful, Θ_1, or eventually Θ_2, must satisfy certain regularity conditions. Limiting ourselves to the case where only Θ_1 is involved, we will assume that the following conditions are satisfied.

(1) At each of its points, the set Θ_1 admits a tangent linear space. More precisely, let ξ

be an arbitrary element of Θ_1. Then there exists a linear affine subspace L_ξ of $\mathcal{E}$ of dimension $s < r$ such that if Π is a projection of $\mathcal{E}$ onto L_ξ there is a function $\eta(u)$ tending to zero with u such that

$$\|(I - \Pi)(\theta - \xi)\| \leqq \eta\{\|\theta - \xi\|\}\|\theta - \xi\| \tag{125}$$

for every $\theta \in \Theta_1$.

(2) The projection Π being as above, for every sequence $\{u_n\}$ in $\mathcal{E}$ tending to ξ there exists a sequence $\{v_n\}$ in Θ_1 such that $\sqrt{n}\|\Pi(u_n - v_n)\|$ tends to zero as n tends to infinity.

Let then $\{T_n\}$ be of the class $\mathfrak{D}$ and let $\{\tau_n\}$ be the associated estimate of section 5. Let $\{\hat{\xi}_n\}$ be a sequence of estimates taking their values in Θ_1 and such that

$$\sqrt{n}(T_n - \hat{\xi}_n)'\Gamma(\tau_n)\sqrt{n}(T_n - \hat{\xi}_n) - \inf_{\xi \in \Theta_1} n(T_n - \xi)'\Gamma(\tau_n)(T_n - \xi) < \epsilon_n \tag{126}$$

for some sequence $\{\epsilon_n\}$ tending to zero as n tends to infinity. Assume that $\xi \in \Theta_1$ is the true value of the parameter and let Π_ξ be the projection on L_ξ orthogonally with respect to $\Gamma(\xi)$. It is easily seen that as n tends to infinity the difference $\sqrt{n}\Pi_\xi T_n - \sqrt{n}\xi_n$ tends to zero in probability. The convergence is uniform on the compacts of Θ_1. From this and the theorems of Slutsky, one concludes that

$$\mathcal{L}\{\sqrt{n}(\hat{\xi}_n - \xi) \mid \xi\} \to \mathcal{N}\{0, \Gamma_\xi \Gamma^{-1}(\xi)\Pi'_\xi\} \tag{127}$$

uniformly in the $\mathcal{T}_c$ sense in the compacts of Θ_1. Similar results apply if $\xi \in \Theta_1$ is considered to be a function of some other parameter $\omega \in \Omega$ provided that the function $\omega \to \xi(\omega)$ be sufficiently regular. Another result of interest for tests of hypotheses is the following. Let $\chi^2(\lambda)$ denote a noncentral χ^2 with $r - s$ degrees of freedom and noncentrality parameter λ.

The set Θ_1 being as above, let

$$Q_n(\theta) = \sqrt{n}(T_n - \theta)'\Gamma(\tau_n)\sqrt{n}(T_n - \theta) - \inf_{\xi \in \Theta_1} \sqrt{n}(T_n - \xi)'\Gamma(\tau_n)\sqrt{n}(T_n - \xi), \tag{128}$$

$$\lambda_n(\theta) = \inf_{\xi \in \Theta_1} \sqrt{n}(\theta - \xi)'\Gamma(\theta)\sqrt{n}(\theta - \xi). \tag{129}$$

Assume that for each n the distribution of T_n corresponds to a value θ_n of the parameter and that the sequence $\{\theta_n\}$ stays in a compact of Θ. Then if $\lambda_n(\theta_n)$ tends to infinity with n the random variables $Q_n(\theta_n)$ tend in probability to infinity. If on the contrary for some $\xi \in \Theta_1$ the sequence $\{\sqrt{n}(\theta_n - \xi)\}$ stays in a compact of $\mathcal{E}$ then

$$\mathcal{L}\{Q_n(\theta_n) \mid \theta_n\} - \mathcal{L}\{\chi^2[\lambda_n(\theta_n)]\} \tag{130}$$

tends to zero in the $\mathcal{T}_c$ sense as n tends to infinity.

In this case let $\gamma_n = \sqrt{n}(I - \Gamma_\xi)(\theta_n - \xi)$. Then one can also say that

$$\mathcal{L}\{Q_n(\theta_n) \mid \theta_n\} - \mathcal{L}\{\chi^2[\gamma'_n\Gamma(\xi)\gamma_n]\} \tag{131}$$

tends to zero as n tends to infinity.

The foregoing discussion should be sufficient to indicate some of the methods available.

We shall only make the further remark that the matrix $\Gamma(\tau_n)$ could be replaced by any one of a variety of matrices without affecting the results. Note however that it cannot always be replaced by $\Gamma(\xi)$, since estimates minimizing $\sqrt{n}(T_n - \xi)' \Gamma(\xi)(T_n - \xi)$ for $\xi \in \Theta_1$ need not be consistent.

8. An application to asymptotically similar tests

The investigation presented in this paper arose naturally from a variety of motives, such as the desire to escape the stringency of Wald's conditions, a need for a link between the B.A.N. estimates and maximum likelihood estimates, etc. However, the decisive factor leading to this study was a desire to encompass in a general body of theory certain results recently published by Neyman in [4]. The situation considered by this author is the following. The parameter space Θ is a two-dimensional half plane, and to each $\theta \in \Theta$ corresponds a density $p(x, \theta)$. Denote by $\theta^{(1)}$ and $\theta^{(2)}$, respectively, the first and second coordinates of the vector θ. The problem is to test the hypothesis $H = \{\theta: \theta^{(2)} = 0\}$ against the alternative $H^c = \{\theta: \theta^{(2)} > 0\}$. Let $\varphi_1[x, \theta^{(1)}, \theta^{(2)}]$ be the partial derivative of $\log p(x, \theta)$ with respect to $\theta^{(1)}$ and let $\varphi_2[u, \theta^{(1)}, \theta^{(2)}]$ be the corresponding derivative with respect to $\theta^{(2)}$, both evaluated at $\theta = [\theta^{(1)}, \theta^{(2)}]$.

Let $f(x, \theta^{(1)})$ be a function such that

$$E\{f[X, \theta^{(1)}] \mid \theta = [\theta^{(1)}, 0]\} = 0 . \tag{132}$$

Let $\rho[\theta^{(1)}]$ be the regression coefficient of $f[X, \theta^{(1)}]$ on $\varphi_1[X, \theta^{(1)}, 0]$ when $\theta = [\theta^{(1)}, 0]$ is the true value and let $s[\theta^{(1)}]$ be the standard deviation of $F[X, \theta^{(1)}] = f[X, \theta^{(1)}] - \rho[\theta^{(1)}]\varphi_1[X, \theta^{(1)}, 0]$ under the same hypothesis on θ.

Under suitable conditions all the preceding quantities exist and furthermore

$$E\{F[X, \theta^{(1)}] \mid \theta = [\theta^{(1)}, 0]\} = 0 . \tag{133}$$

If $\theta^{(1)}$ is known a test of the hypothesis $\theta^{(2)} = 0$ against $\theta^{(2)} > 0$ can be obtained by rejecting the hypothesis if

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} F[x_j, \theta^{(1)}] \geqq k s[\theta^{(1)}] . \tag{134}$$

If $\theta^{(1)}$ is not known, suppose that a sequence $\{\vartheta_n\}$ of estimates of $\theta^{(1)}$ is available and that $\{\sqrt{n} | \vartheta_n - \theta^{(1)} | \}$ is bounded in probability. Then one might think of replacing $\theta^{(1)}$ by its estimated value $\vartheta_n(x_1, \cdots, x_n)$ in the function $F[x, \theta^{(1)}]$ and in $\rho(\theta^{(1)})$ and $s[\theta^{(1)}]$. Neyman shows that under suitable regularity conditions the test so obtained is asymptotically similar in $\theta^{(1)}$.

The conditions given in [4] being somewhat too stringent, it might not be useless to indicate alternative hypotheses under which theorem 5 of [4] is still valid. One such set of hypotheses is the following.

(a) $\log p[x, \theta^{(1)}, 0]$ is twice continuously differentiable in $\theta^{(1)}$ and $f[x, \theta^{(1)}]$ is continuously differentiable in $\theta^{(1)}$.

(b) $E\{f^2[X, \theta^{(1)}] | \theta = [\theta^{(1)}, 0]\}$ and $E\{\varphi_1^2[X, \theta^{(1)}, 0] | \theta = [\theta^{(1)}, 0]\}$ are finite and the second of these expectations is positive.

(c) The integrals

$$\int \varphi_1[x, \theta^{(1)}, 0] p[x, \theta^{(1)}, 0]\, dx \quad \text{and} \quad \int f[x, \theta^{(1)}] p[x, \theta^{(1)}, 0]\, dx \tag{135}$$

are equal to zero and can be differentiated once under the integral sign.

(d) The functions $\rho[\theta^{(1)}]$ and $s[\theta^{(1)}]$ are continuous in $\theta^{(1)}$ and $s[\theta^{(1)}]$ is positive for every $\theta^{(1)}$.

(e) Let $f'[x, \theta^{(1)}]$, $\varphi_1'[x, \theta^{(1)}, 0]$ be the derivatives of $f[x, \theta^{(1)}]$ and $\varphi_1[x, \theta^{(1)}, 0]$ with respect to $\theta^{(1)}$. For every $\theta^{(1)}$ there is an open interval $V[\theta^{(1)}]$ centered at $\theta^{(1)}$ and a function $H[x, \theta^{(1)}]$ such that

$$\text{(136)} \qquad \text{(i)} \qquad \int H[x, \theta^{(1)}]\, p[x, \theta^{(1)}, 0]\, dx < \infty .$$

(ii) For every $t \in V[\theta^{(1)}]$ and every x the following inequalities hold:

$$\text{(137)} \qquad \begin{aligned} |\varphi_1'(x, t, 0)| &\leqq H[x, \theta^{(1)}] \\ |f'(x, t)| &\leqq H[x, \theta^{(1)}] . \end{aligned}$$

It seems reasonable to expect that the "best" test of the family described above is obtainable by taking $f[x, \theta^{(1)}] = \varphi_2[x, \theta^{(1)}, 0]$. That such is the case has been shown by Neyman [11].

To apply the results of the preceding section to the present case suppose that assumptions (A) are satisfied and that $\{T_n\}$ is a sequence of estimates of class $\mathfrak{D}$ constructed from parent estimates $\{t_n\}$ of class $\mathcal{C}$ and having associated estimates $\{\tau_n\}$. Let H be a subset of Θ to be tested against H^c. For any sequence of tests $\{\psi_n\}$ let $\beta_n(\psi_n, \theta)$ be the power of the test ψ_n at θ. It will be convenient to say that a sequence $\{\psi_n\}$ is asymptotically similar of size α uniformly on compacts if $\beta_n(\psi_n, \theta)$ tends to α uniformly on the compacts of H. To simplify we will abridge "asymptotically similar uniformly on compacts" to K.S. Among K.S. sequences of size α there might exist sequences $\{\omega_n\}$ such that for any other K.S. sequence $\{\psi_n\}$ of the same size

$$\text{(138)} \qquad \limsup_{n\to\infty} \sup \{\beta_n(\psi_n, \theta) - \beta_n(\omega_n, \theta);\ \theta \in K \cap H^c\} \leqq 0$$

for every compact $K \subset \Theta$. Such a sequence would be called asymptotically uniformly most powerful on compacts among K.S. sequences and, for short, K.U.M.P.S.

To find such a sequence it is most convenient to consider first a simpler problem.

Let $\{\theta_n\}$ and $\{\theta_n^*\}$ be two arbitrary sequences in Θ converging to the same point $\theta_0 \in \Theta$. Let $\xi_n = \sqrt{n}(\theta_n - \theta_0)$ and $\xi_n^* = \sqrt{n}(\theta_n^* - \theta_0)$ and finally $X_n = \sqrt{n}(T_n - \theta_0)$. Furthermore, let $f_n(x, \xi_n)$ and $g_n(x, \xi_n)$ be defined by

$$\text{(139)} \qquad f_n(x, \xi_n) = \frac{[\det \Gamma(\theta_0)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(x - \xi_n)' \Gamma(\theta_0)(x - \xi_n)\}$$

and

$$\text{(140)} \qquad g_n(x, \xi_n) = \frac{[\det \Gamma(\theta_n)]^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}(x - \xi_n)' \Gamma(\theta_n)(x - \xi_n)\}.$$

It is easily seen that, μ_n being the measure defined in theorem 2, section 6, the quantity

$$\text{(141)} \qquad \int |g_n(x, \xi_n) - f_n(x, \xi_n)|\, d\mu_n$$

tends to zero as n tends to infinity. Therefore, for a sequence $\{\theta_n\}$ tending to θ_0 one can approximate the measures $\bar{Q}_{\theta_n, n}$ of theorem 2 by the simpler measures $\hat{Q}_{\xi_n, n}$ defined by

$$\text{(142)} \qquad \hat{Q}_{\xi_n, n}(S) = \int_S f_n(x, \xi_n)\, d\mu_n .$$

The same proposition holds true *mutatis mutandis* for the sequence $\{\theta_n^*\}$.

Instead of trying to find a test of $P_{\theta_n}^{(n)}$ against $P_{\theta_n^*}^{(n)}$ one may as well try to find a test of $\hat{Q}_{\xi_n, n}$ against $\hat{Q}_{\xi_n^*, n}$, the results so obtained being asymptotically equivalent according to theorems 1 and 2.

According to the fundamental lemma of the Neyman-Pearson theory, for any level of significance α there is a constant $k_n(\alpha, \xi_n, \xi_n^*)$ such that the best test at level α of the hypothesis $\hat{Q}_{\xi_n, n}$ against $\hat{Q}_{\xi_n^*, n}$ is given by

$$(143) \qquad \bar{\psi}_n(X_n) = \begin{cases} 0 & \text{if} \quad (\xi_n^* - \xi_n)' \Gamma(\theta_0)\left(X_n - \frac{\xi_n + \xi_n^*}{2}\right) > k_n \\ 1 & \text{if} \quad (\xi_n^* - \xi_n)' \Gamma(\theta_0)\left(X_n - \frac{\xi_n + \xi_n^*}{2}\right) < k_n \end{cases}$$

and some number between zero and unity if

$$(144) \qquad (\xi_n^* - \xi_n)' \Gamma(\theta_0) \left[X_n - \left(\frac{\xi_n + \xi_n^*}{2}\right)\right] = k_n .$$

Let h be a number such that

$$(145) \qquad \frac{1}{2\pi} \int_h^\infty e^{-t^2/2} dt = \alpha$$

and let δ_n be the difference $\xi_n^* - \xi_n$. Simple algebra shows that, since

$$(146) \qquad \mathcal{L}\{X_n - \xi_n \mid \xi_n\} \to \mathcal{N}[0, \Gamma^{-1}(\theta_0)],$$

the tests $\{\psi_n\}$ obtained by rejecting ξ_n if

$$(147) \qquad \delta_n' \Gamma(\theta_0)(x - \xi_n) \geqq h\sqrt{\xi_n' \Gamma(\theta_0)\delta_n}$$

are most powerful for testing ξ_n against ξ_n^* and have levels of significance tending to α.

To come back to the problem of finding asymptotically similar tests, assume that the hypothesis specifies that θ lies in a hyperplane H of $\mathcal{E}$ and that Θ is the part of $\mathcal{E}$ on one side of this hyperplane. Take $\theta_0 \in H$ as origin of coordinates and let e be an arbitrary vector not contained in H. Every element x of $\mathcal{E}$ can be written in a unique manner as $x = u + ve$, with $u \in H$. It will be assumed that for $x \in H^c$ the number v is positive. Let Π_0 be the projection on H orthogonally with respect to $\Gamma(\theta_0)$. Then every $x \in \mathcal{E}$ can also be written

$$(148) \qquad x = \Pi_0 x + (I - \Pi_0)x = u + v\Pi_0 e + v(I - \Pi_0)e .$$

Let $e_0 = (I - \Pi_0)e$ and $z = u + v\Pi_0 e$. Let $\{\theta_n^*\}$ be a sequence of elements of Θ tending to θ_0. Let $\xi_n^* = \sqrt{n}(\theta_n^* - \theta_0)$ and let $\xi_n^* = \xi_n + \eta_n e_0$ with $\xi_n = \Pi_0 \xi_n^*$.

For testing the hypothesis $\xi = \xi_n$ against $\xi = \xi_n^*$ the test derived above reduces to the test ψ_n obtained by rejecting the hypothesis if

$$(149) \qquad \eta_n v[e_0' \Gamma(\theta_0) e_0] \geqq h\sqrt{\eta_n^2 e_0' \Gamma(\theta_0) e_0} ,$$

that is, since η_n is positive

$$(150) \qquad v \geqq h[e_0' \Gamma(\theta_0) e_0]^{-1/2} .$$

Such a test depends on the particular sequence $\{\theta_n^*\}$ only by the occurrence of the limit point θ_0 in $\Gamma(\theta_0)$ and consequently e_0. However, for any sequence $\{\theta_n^*\}$ tending to θ_0 the matrix $\Gamma(\tau_n)$ tends in probability to $\Gamma(\theta_0)$ so that by Slutsky's theorem the tests $\{\omega_n\}$ obtained by rejecting H if

$$v \geqq h\,[e'(I-\Pi_{\tau_n})'\,\Gamma(\tau_n)(I-\Pi_{\tau_n})\,e]^{-1/2} \tag{151}$$

are asymptotically equivalent to $\{\psi_n\}$ for every sequence $\{\theta_n^*\}$ tending to θ_0.

The above argument applies to $\{\omega_n\}$ whatever may be the limit point $\theta_0 \in H$. Therefore, $\{\omega_n\}$ is a K.S. sequence of size α for testing H against H^c.

It remains to show that this sequence is K.U.M.P.S. If this were not true, there would exist a sequence $\{\theta_n^*\}$ contained in a compact, an $\epsilon > 0$ and a K.S. sequence $\{\bar{\psi}_n\}$ of size α such that

$$\limsup_{n\to\infty} \beta_n(\psi_n, \theta_n^*) - \beta_n(\omega_n, \theta_n^*) \geqq \epsilon > 0\,. \tag{152}$$

By extracting a subsequence, if necessary, one can assume that $\{\theta_n^*\}$ converges to some point $\theta_0 \in \Theta$. However, if $\theta_0 \in H^c$ or if $\sqrt{n}(\theta_n^* - \theta_0) = \xi_n^* = \xi_n + \eta_n e_0$ with $\theta_0 \in H$ and $\{\eta_n\}$ unbounded then $\lim\sup \beta_n(\omega_n, \theta_n^*) = 1$ so that the preceding inequality is impossible. Hence $\{\eta_n\}$ must stay bounded. But then $\{\bar{\psi}_n\}$ is asymptotically of size α for testing ξ_n against ξ_n^* and therefore cannot be asymptotically more powerful than $\{\omega_n\}$. This completes the proof of the proposition: $\{\omega_n\}$ is K.U.M.P.S. of size α for testing H against H^c.

For the situation considered earlier where Θ is a two-dimensional half plane $\Theta = \{\theta\colon \theta^{(2)} \geqq 0\}$ and H is the line $H = \{\theta\colon \theta^{(2)} = 0\}$, the sequence of tests $\{\omega_n\}$ takes a particularly simple form. Let $\gamma_{ij}(t)$ be the elements of the matrix $\Gamma(t)$ and let $\Delta(t)$ be the determinant of $I(t)$, that is, $\Delta(t) = \gamma_{11}(t)\gamma_{22}(t) - \gamma_{12}^2(t)$. Let $T_n^{(2)}$ be the second coordinate of the vector T_n. The critical region of the test ω_n is given by the inequality

$$\sqrt{n}\,T_n^{(2)} \geqq h\sqrt{\frac{\gamma_{11}(\tau_n)}{\Delta(\tau_n)}}\,. \tag{153}$$

For any vector $t = [t^{(1)}, t^{(2)}]$ let

$$\begin{aligned} U_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varphi_1[x, t^{(1)}, t^{(2)}] \\ V_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varphi_2[x, t^{(1)}, t^{(2)}]\,. \end{aligned} \tag{154}$$

By definition T_n is derived from a parent estimate t_n by the formula

$$T_n = t_n + \Gamma^{-1}(t_n)\,A_n'(t_n) = t_n + \frac{1}{\sqrt{n}}\,\Gamma^{-1}(t_n)\begin{pmatrix} U_n(t_n) \\ V_n(t_n)\end{pmatrix} \tag{155}$$

giving

$$T_n^{(2)} = t_n^{(2)} + \frac{1}{\Delta_n(t_n)\sqrt{n}}\,[\gamma_{11}(t_n)\,V_n(t_n) - \gamma_{12}(t_n)\,U_n(t_n)]\,. \tag{156}$$

If $t_n^{(2)}$ is large the formula giving the test ω_n differs noticeably from the ones proposed by Neyman. However, if $t_n^{(2)}$ is small the situation is quite different. Let s_n be the vector of coordinates $s_n^{(1)} = t_n^{(1)}$ and $s_n^{(2)} = 0$. According to assumptions (A)

$$A_n'(t_n) = A_n'(s_n) - B_n(s_n, t_n)(t_n - s_n) \tag{157}$$

where $B_n(s_n, t_n)$ is an average of $B_n(t)$ between s_n and t_n. Denote the elements of $B_n(s_n, t_n)$ by $\bar{\gamma}_{ij}(n)$. Expanding U_n and V_n by Taylor's formula one obtains

$$(158) \qquad \sqrt{n}\, T_n^{(2)} = \frac{1}{\Delta(t_n)} [\gamma_{11}(t_n) V_n(s_n) - \gamma_{12}(t_n) U_n(s_n)] + R_n$$

with

$$(159) \qquad R_n = \sqrt{n}\, t_n^{(2)} \left\{ 1 - \frac{\gamma_{11}(t_n)\bar{\gamma}_{22}(n) - \gamma_{12}(t_n)\bar{\gamma}_{12}(n)}{\Delta(t_n)} \right\}.$$

Consequently, if $\{\theta_n\}$ is a sequence of values of θ such that for some $\theta_0 \in H$ the vectors $\sqrt{n}(\theta_n - \theta_0)$ stay bounded, then for every $\epsilon > 0$ the quantity $P\{R_n > \epsilon | \theta_n\}$ tends to zero as n tends to infinity. Consider then a sequence $\{\psi_n\}$ of tests defined by the critical regions

$$(160) \qquad \frac{1}{\Delta(s_n)} [\gamma_{11}(s_n) V_n(s_n) - \gamma_{12}(s_n) U_n(s_n)] \geqq h \sqrt{\frac{\gamma_{11}(s_n)}{\Delta(s_n)}}.$$

Slutsky's theorem implies again that if for some $\theta_0 \in H$ the vectors $\sqrt{n}(\theta_n - \theta_0)$ stay bounded then

$$(161) \qquad \lim_{n\to\infty} \beta_n(\psi_n, \theta_n) - \beta_n(\omega_n, \theta_n) = 0.$$

One can also verify easily that when $\{\theta_n\}$ tends to $\theta_0 \in H$ but $\sqrt{n}\theta_n^{(2)}$ tends to infinity, the power $\beta_n(\psi_n, \theta_n)$ tends to unity. Consequently, one can say that for every sequence $\{\theta_n\}$ converging to a point of H

$$(162) \qquad \lim_{n\to\infty} \beta_n(\psi_n, \theta_n) - \beta_n(\omega_n, \theta_n) = 0.$$

The sequence of tests $\{\psi_n\}$ is a member of the class proposed by Neyman. More precisely $\{\psi_n\}$ is exactly the sequence of tests that Neyman has shown to be the "best" of the proposed class. The following proposition summarizes the situation.

Let assumptions (A) be satisfied and let Θ be the half plane $\Theta = \{\theta: \theta^{(2)} \geqq 0\}$. Let H be the hypothesis $H = \{\theta: \theta^{(2)} = 0\}$. Let $\{T_n\}$ be a sequence of estimates of the class $\mathfrak{D}$ with parent estimates $\{t_n\}$. Let s_n be the vector of coordinates $s_n^{(1)} = t_n^{(1)}$ and $s_n^{(2)} = 0$. Let ω_n be the test whose critical region is given by

$$(163) \qquad \sqrt{n}\, T_n^{(2)} \geqq h \sqrt{\frac{\gamma_{11}(t_n)}{\Delta(t_n)}}.$$

Let ψ_n be the test whose critical region is given by

$$(164) \qquad \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \varphi_2[x_j, t_n^{(1)}, 0] - \frac{\gamma_{12}(s_n)}{\gamma_{11}(s_n)} \varphi_1[x_j, t_n^{(1)}, 0] \right\} \geqq h \sqrt{\frac{\Delta(s_n)}{\gamma_{11}(s_n)}}.$$

PROPOSITION 2. *Both sequences $\{\omega_n\}$ and $\{\psi_n\}$ are K.S. sequences of size α. The sequence $\{\omega_n\}$ is K.U.M.P.S. For every sequence θ_n tending to a point of H*

$$(165) \qquad \lim_{n\to\infty} \beta_n(\psi_n, \theta_n) - \beta_n(\omega_n, \theta_n) = 0.$$

A necessary and sufficient condition that $\{\psi_n\}$ be K.U.M.P.S. is that $\beta_n(\psi_n, \theta)$ converge to unity uniformly on the compacts of H^c.

Since the tests ψ_n are very much simpler to use than the tests ω_n one might try to replace ω_n by ψ_n whenever possible. Note that according to the preceding propositions there is no substantial difference in their performance close to the hypothesis H. However, to obtain this proposition we have assumed that the estimate available $\{t_n\}$ possesses the strong properties of consistency of the class $\mathcal{C}$ not only under the hypothesis but under the alternatives as well. It is easily seen that an estimate $\{t_n\}$ which is of the class $\mathcal{C}$ only for the hypothesis H will lead to K.S. tests but not necessarily to K.U.M.P.S. tests. Even if the sequence $\{t_n\}$ is of the class $\mathcal{C}$ for the whole of Θ, the tests ω_n and ψ_n might differ essentially for values of θ remote from H.

To the author's knowledge, no simple condition to insure that $\beta_n(\psi_n, \theta)$ converges to unity uniformly on the compacts of H^c has yet been found.

REFERENCES

[1] J. NEYMAN, "Contributions to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1949, pp. 239–273.

[2] E. W. BARANKIN and J. GURLAND, "On asymptotically normal, efficient estimators: I," *Univ. of California Publ. in Stat.*, Vol. 1 (1951), pp. 89–130.

[3] A. WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[4] J. NEYMAN, "Sur une famille de tests asymptotiques des hypothèses statistiques composées," *Trabajos de Estadística*, Vol. 5 (1954), pp. 161–168.

[5] E. SLUTSKY, "Ueber stochastiche asymptotem und grenzwerte," *Metron*, Vol. 5 (1925), No. 3, pp. 3–90.

[6] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[7] C. KRAFT and L. LE CAM, "A remark on the roots of the maximum likelihood equation," to be published.

[8] E. PARZEN, "On uniform convergence of families of sequences of random variables," *Univ. of California Publ. in Stat.*, Vol. 2 (1954), pp. 23–54.

[9] L. LE CAM, "On the asymptotic properties of Bayes solutions," to be published.

[10] ———, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *Univ. of California Publ. in Stat.*, Vol. 1 (1953), pp. 277–330.

[11] J. NEYMAN, personal communication.