ON THE ASYMPTOTIC EFFICIENCY OF TESTS AND ESTIMATES

By R. R. BAHADUR

Indian Statistical Institute, Calcutta

SUMMARY. Let $x_1, x_2, \ldots$ be a sequence of independent and identically distributed observations with distribution determined by a real valued parameter θ. For each $n = 1, 2, \ldots$, let $T_n = T_n(x_1, x_2, \ldots, x_n)$ be a statistic such that the sequence $\{T_n\}$ is a consistent estimate of θ. It is shown, under weak regularity conditions on the sample space of a single observation, that the asymptotic effective standard deviation of T_n cannot be less than $[nI(\theta)]^{-\frac{1}{2}}$. The asymptotic effective standard deviation of T_n is defined, roughly speaking, as the solution τ of the equation $P(|T_n - \theta| > \epsilon|\theta) = P(|N| > \epsilon/\tau)$ when n is large and ϵ is a small positive number, where N denotes a standard normal variable. It is also shown, under stronger regularity conditions, that the asymptotic effective standard deviation of the maximum likelihood estimate of θ is $[nI(\theta)]^{-\frac{1}{2}}$. These conclusions concerning estimates are derived from certain conclusions concerning the relative efficiency of alternative statistical tests based on large samples.

1. Introduction

Let X be an abstract sample space of points x, and suppose that the distribution of x is determined by a (not necessarily real valued) parameter θ taking values in a set Θ. Let g be a given real valued function of θ and suppose that it is required to estimate g. For each $n = 1, 2, \ldots$ let $(x_1, x_2, \ldots, x_n) = x_{(n)}$ (say) denote n independent observations on x, and let $X_{(n)}$ denote the sample space of $x_{(n)}$. An *estimate* is defined (without explicit reference to g) to be a sequence $\{T_n\} = \mathbf{T}$ (say), such that $T_n = T_n(x_{(n)})$ is a real valued measurable function on $X_{(n)}$, $(n = 1, 2, \ldots)$. An estimate $\mathbf{T} = \{T_n\}$ is said to be consistent if for each θ and each $\epsilon > 0$, $P(|T_n - g(\theta)| \geqslant \epsilon|\theta) \to 0$ as $n \to \infty$; $\mathbf{T}$ is consistent and asymptotically normal (c.a.n., in short) if for each θ in Θ there exists a sequence $\{\sigma_n\}$ of positive real numbers such that $\lim_{n\to\infty} \sigma_n(\theta) = 0$ and $\lim_{n\to\infty} P([T_n - g(\theta)]/\sigma_n(\theta) \leqslant x|\theta) = P(N \leqslant x)$ for every x, where N denotes a normally distributed random variable with $E(N) = 0$ and $E(N^2) = 1$. $\sigma_n^2(\theta)$ is then called the asymptotic variance of T_n when θ obtains.

The classical theory of estimation from large samples (cf., e.g., Fisher (1922, 1925), Neyman (1949), Gurland and Barankin (1951), LeCam (1953), Kallianpur and Rao (1955); cf. also the references cited by these authors) has been concerned mainly with c.a.n. estimates, the usual criterion of assessment of a particular estimate being its asymptotic variance. As is well known, rigorous theoretical development of the criterion just mentioned has proved full of complications and difficulties. One of the reasons for the difficulties encountered is surely that the asymptotic variance σ_n^2 of a statistic T_n has a very weak (or at least ill-determined) relation to the actual concentration of the distribution of T_n at the value of g —unless , of course, T_n happens to be exactly normally distributed with mean g, and σ_n^2 is the actual variance of T_n.

To put it in another way, in comparing estimates which are c.a.n., but otherwise arbitrary, comparisons of their asymptotic variances appear to lack justification. Indeed, Basu (1956) has given an example where such comparisons seem definitely misleading; this example exhibits two c.a.n. estimates, $\{T_n\}$ and $\{U_n\}$, with asymptotic variances $\{a_n\}$ and $\{b_n\}$ respectively, such that $a_n/b_n \to 0$ as $n \to \infty$ but, for every $\epsilon > 0$,

$$\frac{P(|T_n - g(\theta)| \geqslant \epsilon|\theta)}{P(|U_n - g(\theta)| \geqslant \epsilon|\theta)} \to \infty. \quad \ldots \quad (1.1)$$

One of the objects of this paper is to suggest an approach, parallel to the classical theory, in which a given c.a.n. estimate $\{T_n\}$ (or indeed any estimate) is discussed in terms of 'effective standard deviations', the latter being scaling constants which are well articulated with the actual probability distributions of the estimating statistics T_n. This approach can be effected in several different ways. The particular version presented here is tentative, and has no claim to logical necessity.

Consider a c.a.n. estimate $\mathbf{T} = \{T_n\}$, and suppose that for a particular n, θ, and $\epsilon > 0$ we wish, for some practical or theoretical reason, to compute $P(|T_n - g| \geqslant \epsilon)$. If σ_n^2 is the asymptotic variance of T_n, $P(|N| \geqslant \epsilon/\sigma_n)$ is an approximation to the probability required. This approximation is, however, of unknown accuracy, and that is precisely why σ_n^2 is in general an unsatisfactory index of the performance of T_n. Suppose we define τ as follows:

Definition 1.1 : For any real valued statistic T_n, any θ, and any $\epsilon > 0$, $\tau = \tau_g(T_n, \epsilon, \theta)$ is the solution of the equation

$$P(|N| \geqslant \epsilon/\tau) = P(|T_n - g(\theta)| \geqslant \epsilon|\theta), \qquad (0 \leqslant \tau \leqslant \infty). \quad \ldots \quad (1.2)$$

Then τ achieves what σ is supposed to do, i.e., the right side of (1.2) can be computed exactly by entering a standard table of the normal distribution with ϵ/τ. Consequently, τ might be called the effective standard deviation of T_n when (a) T_n is regarded as a point estimate of g, (b) θ obtains, and (c) it is required to compute the right side of (1.2). If T_n is exactly normally distributed with mean g then, for every ϵ, $\tau_g(T_n, \epsilon, \theta)$ equals, as it should, the actual standard deviation of T_n.

Although suggested by the study of c.a.n. estimates, Definition 1.1 is applicable to any estimate whatsoever. In particular, if T_n and U_n are any two real valued statistics then $P(|T_n - g(\theta)| \geqslant \epsilon|\theta) \geqslant P(|U_n - g(\theta)| \geqslant \epsilon|\theta)$ if and only if $\tau_g(T_n, \epsilon, \theta) \geqslant \tau_g(U_n, \epsilon, \theta)$. Again, $\mathbf{T} = \{T_n\}$ is a consistent estimate of g if and only if, for each ϵ and each θ in Θ, $\tau_g(T_n, \epsilon, \theta) \to 0$ as $n \to \infty$. In the following development based on Definition 1.1, we consider the class of all consistent estimates. The restriction to c.a.n. estimates is not made henceforth because it is unnecessary, and because the present definitions and conclusions concerning the wider class of estimates may have some bearing on formulations of large sample estimation theory (e.g. Savage (1954)) in which asymptotic normality is not an *a priori* requirement.

Let $\mathbf{T} = \{T_n\}$ and $\mathbf{U} = \{U_n\}$ be two estimates of g. We shall define the upper asymptotic efficiency of $\mathbf{T}$ relative to $\mathbf{U}$ when θ obtains, $\bar{e}_g(\mathbf{T}, \mathbf{U}|\theta)$ say, as follows :

$$\bar{e}_g(\mathbf{T}, \mathbf{U}|\theta) = \overline{\lim_{\epsilon\to 0}}\ \overline{\lim_{n\to\infty}}\ \{\tau_g^2(U_n, \epsilon, \theta)/\tau_g^2(T_n, \epsilon, \theta)\}. \qquad \dots \quad (1.3)$$

Similarly, the lower asymptotic relative efficiency, $\underline{e}_g(\mathbf{T}, \mathbf{U}|\theta)$ say, is defined by (1.3) but with $\overline{\lim}_{\epsilon\to 0}\ \overline{\lim}_{n\to\infty}$ replaced by $\underline{\lim}_{\epsilon\to 0}\ \underline{\lim}_{n\to\infty}$. In the brackets on the right side of (1.3), and elsewhere in this paper, we take the ratios 0/0 and ∞/∞, if they occur, to be equal to 1. Then $\underline{e}$ and $\bar{e}$ are always well defined, with $0 \leqslant \underline{e} \leqslant \bar{e} \leqslant \infty$.

The following considerations are relevant to the definitions of the preceding paragraph. In the first place, for given T_n, ϵ, and θ, τ^2 is in many cases of the order $1/n$, so that $\lim_{n\to\infty} \{\tau_g^2(U_n, \epsilon, \theta)/\tau_g^2(T_n, \epsilon, \theta)\}$ exists in such cases, and (roughly speaking) equals the limiting ratio of sample sizes required to obtain comparable probabilities of the event $\{|\text{estimated value—actual value}| \geqslant \epsilon\} = E$ (say). In the second place, when n is large, the distribution of any tolerable estimating statistic will be concentrated in a neighbourhood of g, so that small values of ϵ become of primary interest. Since n is made infinite as the first step in the definition of e, it is therefore appropriate that ϵ then be made to tend to zero.

It may be added here that if we were to let $n \to \infty$ and $\epsilon \to 0$ simultaneously, by setting $\epsilon = \lambda/\sqrt{n}$ where λ is a constant, then the above definition of relative efficiency would reduce to the classical definition for c.a.n. estimates with asymptotic variances proportional to $1/n$. (cf. remark 3 in Section 6).

It is of some interest and importance to the present formulation that if $\{T_n\}$ is any consistent estimate of g then, for large n, τ^2 is approximately inversely proportional to log $[1/P(E)]$; more precisely, for fixed $\epsilon > 0$ and θ,

$$\frac{2}{\epsilon^2} \log P(|T_n - g(\theta)| \geqslant \epsilon|\theta) = -\frac{1}{\tau_g^2(T_n, \epsilon, \theta)}[1+o(1)] \qquad \dots \quad (1.4)$$

as $n \to \infty$, where τ is given by (1.2). This asymptotic relation is an immediate consequence of Lemma 2.3. It follows from (1.4) that definitions and conclusions concerning effective variances τ^2 can be readily phrased in terms of probabilities of deviations instead.

It follows from (1.4), in particular, that if $\mathbf{T} = \{T_n\}$ and $\mathbf{U} = \{U_n\}$ are consistent estimates of g, and $\bar{e}(\mathbf{T}, \mathbf{U}|\theta) < 1$, then, for each sufficiently small $\epsilon > 0$, $P(|T_n - g| \geqslant \epsilon|\theta) > P(|U_n - g| \geqslant \epsilon|\theta)$ for all sufficiently large n; in fact (1.1) holds as $n \to \infty$.

Definition 1.2 : An estimate $\mathbf{T}^* = \{T_n^*\}$ is an asymptotically efficient estimate of g if $\mathbf{T}^*$ is consistent and if, for each θ in Θ, $\bar{e}_g(\mathbf{T}, \mathbf{T}^*|\theta) \leqslant 1$ for all other consistent estimates $\mathbf{T}$.

It is shown in Section 5 that asymptotically efficient estimates exist under fairly general conditions, and that the method of maximum likelihood (m.l.) typically leads to such estimates. The main results may be described as follows. Suppose for simplicity that θ is real valued, that Θ is an interval, and that g is a monotonic and differentiable function with derivative $g'(\theta) \neq 0$. Suppose also that the sample space of a single observation x satisfies certain weak regularity conditions. It is shown that for any consistent $\mathbf{T} = \{T_n\}$ and any θ we then have

$$\underline{\lim}_{\epsilon \to 0} \; \underline{\lim}_{n \to \infty} \; \{n \, \tau_g^2(T_n, \epsilon, \theta)\} \geqslant [g'(\theta)]^2 / I(\theta). \qquad \ldots \quad (1.5)$$

Here $I(\theta)$ corresponds to (and, under certain additional conditions, coincides with) the classical 'information concerning θ contained in x'. It is also shown, under certain additional regularity conditions, that if $\hat{\mathbf{U}} = \{\hat{U}_n\}$ is the m.l. estimate of g then, for each θ,

$$\overline{\lim}_{\epsilon \to 0} \; \overline{\lim}_{n \to \infty} \; \{n \, \tau_g^2(\hat{U}_n, \epsilon, \theta)\} \leqslant [g'(\theta)]^2 / I(\theta). \qquad \ldots \quad (1.6)$$

It follows easily from (1.5) and (1.6) that the m.l. estimate of g is efficient according to Definition 1.2.

Definition 1.3: Let $\mathbf{T} = \{T_n\}$ be an estimate of g, and suppose that there exists a sequence $\{v_n(\theta)\}$ of real numbers ($0 \leqslant v_n \leqslant \infty$ for each n) such that

$$\underline{\lim}_{\epsilon \to 0} \; \underline{\lim}_{n \to \infty} \; \{\tau_g^2(T_n, \epsilon, \theta)/v_n(\theta)\} = 1 = \overline{\lim}_{\epsilon \to 0} \; \overline{\lim}_{n \to \infty} \; \{\tau_g^2(T_n, \epsilon, \theta)/v_n(\theta)\}. \quad \ldots \quad (1.7)$$

Then $v_n(\theta)$ is called the asymptotic effective variance of T_n when θ obtains.

It follows from (1.5), with T_n replaced by $\hat{U}_n$, and (1.6) that $[g'(\theta)]^2/nI(\theta)$ is the asymptotic effective variance of the m.l. estimate of g. In view of (1.4), this conclusion can also be stated as follows : With $\{\hat{U}_n\}$ the m.l. estimate of g,

$$P(|\hat{U}_n - g(\theta)| \geqslant \epsilon | \theta) = \exp\left(-\tfrac{1}{2} n \, \epsilon^2 I_g \, [1 + \delta_n(\epsilon, \theta)]\right) \qquad \ldots \quad (1.8)$$

where

$$I_g(\theta) = I(\theta)/[g'(\theta)]^2 \quad \text{and} \quad \underline{\lim}_{\epsilon \to 0} \; \overline{\lim}_{n \to \infty} \; \{|\delta_n(\epsilon, \theta)|\} = 0. \qquad \ldots \quad (1.9)$$

In the following Section 2 some purely analytical lemmas are stated, the main one (Lemma 2.1) being a result of Chernoff (1952) concerning the distribution of a sum of independent and identically distributed random variables. Lemma 2.1, together with the fundamental lemma of Neyman and Pearson, are used in Section 3 to study the asymptotic efficiency of tests of a simple hypothesis against a simple alternative.

It is shown that when the sample size n is large, and the probability of an error of type two is bounded away from 1, the minimum attainable probability of an error of type one is approximately $\exp(-nH)$, where H is one of the information functions introduced by Kullback and Liebler (1951). (cf. also Savage (1954)). An application of this conclusion to the stochastic comparison of tests (Bahadur, 1960b) is given in Section 4. The theorem of Section 3 is shown in Section 5 to lead to the asymptotic inequality (1.5), by using the fact that a consistent estimate of g also provides a consistent test of the value of g. The inequality (1.6) is established, however, by methods typical of estimation theory.

Connections between asymptotic theories of estimation and of testing hypotheses have appeared in the literature from time to time, in more or less concrete forms. The present paper provides an example of a very concrete and explicit connection. Estimation theory appears here as a limiting case of the theory of tests, with the considerations of Section 4 marking the transition from one theory to the other. In particular, the asymptotic efficiency of the m.l. estimate of θ is formally equivalent to certain conclusions to the effect that tests of $\theta = \theta_0$ based on the m.l. estimate are asymptotically efficient against alternatives θ in the neighbourhood of θ_0.

2. Lemmas

Let $z_1, z_2, \ldots$ be a sequence of independent and identically distributed real valued random variables. For each $n = 1, 2, \ldots$ let

$$S_n = \sum_{i=1}^{n} z_i . \qquad \ldots \quad (2.1)$$

Let $k_1, k_2, \ldots$ be a sequence of constants, and define

$$\alpha_n = P(S_n \geqslant k_n). \qquad \ldots \quad (2.2)$$

Suppose that the following conditions (i)–(iv) are satisfied; these conditions imply, among other things, that $\alpha_n \neq 0$ but $\alpha_n \to 0$ very rapidly as $n \to \infty$. The conditions are (i) that

$$\lim_{n \to \infty} \{k_n/n\} = H \text{ (say), where } -\infty < H < \infty. \qquad \ldots \quad (2.3)$$

Let ϕ denote the moment generating function (m.g.f.) of z, i.e. $\phi(t) = E(e^{tz})$. We suppose (ii) that $\phi(t_1) < \infty$ for some $t_1 > 0$. It then follows from well-known properties of m.g.fs that the set $\{t : \phi(t) < \infty\}$, T say, is an interval which includes positive values. It also follows that ϕ possesses derivatives of all orders in the interior of T, and that the derivatives may be obtained by differentiating $E(e^{tz})$ under the expectation sign. We suppose (iii) that there exists a positive t_0 in the interior of T, such that

$$\frac{\phi'(t_0)}{\phi(t_0)} = H. \qquad \ldots \quad (2.4)$$

Finally, it is assumed (iv) that z is non-degenerate, i.e.

$$P(z = c) < 1 \text{ for every constant } c. \qquad \dots \quad (2.5)$$

The reader may verify that conditions (ii), (iii) and (iv) are satisfied, in particular, if $\phi(t) < \infty$ for all t, $H > E(z)$ and $P(z > H) > 0$.

Define

$$\psi(t) = e^{-tH}\phi(t) = e^{-tH}E(e^{tz}). \qquad \dots \quad (2.6)$$

It follows easily from (2.4) and (2.5) that $\psi(t_0) = \rho$ (say) is the minimum value of ψ, and that $0 < \rho < 1$.

The following lemma, due to Chernoff, is basic to this paper.

Lemma 2.1 : *Suppose that*

$$k_n/n = H \text{ for every } n. \qquad \dots \quad (2.7)$$

Then (a) $\alpha_n \leqslant \rho^n \leqslant [\psi(t)]^n$ *for every* n *and* t, *and* (b) *for any given* ϵ, *with* $0 < \epsilon < \rho$, $\alpha_n \geqslant (\rho-\epsilon)^n$ *for all sufficiently large* n.

Proof : For the proof see Chernoff (1952). An alternative proof, and certain refinements of Lemma 2.1, have been given by Bahadur and Ranga Rao (1960).

It follows from Lemma 2.1 that if (2.7) holds then $n^{-1} \log \alpha_n = \log \rho + o(1)$. The following lemma states that this asymptotic formula remains valid when (2.7) is relaxed to (2.3).

Lemma 2.2 : $\lim\limits_{n \to \infty} \{n^{-1} \log \alpha_n\} = \log \rho$.

Proof: It follows from (2.5) that the second derivatives of $\log \phi(t)$ is positive throughout the interior of T, so that $\phi'(t)/\phi(t)$ is strictly increasing and continuous therein. Consequently, by (2.4), the equation $\phi'(t)/\phi(t) = u$ has a positive solution, $f(u)$ say, in the interior of T, provided $|u-H|$ is sufficiently small. Choose and fix a $u > H$ so that $u-H$ is sufficiently small. Then we have $\alpha_n \geqslant P(S_n \geqslant nu)$ for all sufficiently large n, by (2.2) and (2.3). Hence $\underline{\lim}\ n^{-1} \log \alpha_n \geqslant -u f(u) + \log \phi(f(u))$, by an application of Lemma 2.1. Since $f(u)$ is a continuous function of u in a neighbourhood of $u = H$, and since $\phi(t)$ is continuous in t in a neighbourhood of $t = t_0 = f(H)$, it follows by letting $u \to H$ that $\underline{\lim}\ n^{-1} \log \alpha_n \geqslant -Ht_0 + \log \phi(t_0) = \log \rho$. By taking $u < H$, a similar argument shows that $\overline{\lim}\ n^{-1} \log \alpha_n \leqslant \log \rho$. Thus $\lim\ n^{-1} \log \alpha_n = \log \rho$, and this completes the proof.

Lemma 2.3: $\log P(|N| \geqslant x) = -\dfrac{x^2}{2}[1+o(1)]$ *as* $x \to \infty$.

Proof : This is an immediate consequence of the lemma given in Feller (1957, p. 166).

The next and final lemma is to the effect that the asymptotic effective variance (cf. Section 1) of the sample mean of n independent and identically distributed random variable $z_1, z_2, \dots$ exists and equals n^{-1} times the actual variance of each z. It is assumed now that the m.g.f. of z exists, i.e. that there exists a $\delta > 0$, such that $\phi(t) = E(e^{tz}) < \infty$ for all t with $|t| < \delta$. We also suppose, without additional loss of generality, that

$$E(z) = 0, \text{ and } 0 < E(z^2) < \infty. \qquad \dots \quad (2.8)$$

S_n is defined by (2.1) for each n.

Lemma 2.4: *If* $\lambda_n(\epsilon)$ *is defined by*

$$P(|S_n/n| \geqslant \epsilon) = P(|N| \geqslant \epsilon/\lambda_n(\epsilon)) \qquad \dots \quad (2.9)$$

for $\epsilon > 0$ *and* $n = 1, 2, \dots$ *then*

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \{n \lambda_n^2(\epsilon)\} = E(z^2). \qquad \dots \quad (2.10)$$

Equivalently,

$$P(|S_n/n| \geqslant \epsilon) = e^{-\frac{n\epsilon^2}{2E(z^2)}[1+\delta_n(\epsilon)]} \qquad \dots \quad (2.11)$$

where

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \{\delta_n(\epsilon)\} = 0. \qquad \dots \quad (2.12)$$

Proof: Consider a fixed $\epsilon > 0$. Let t_1 and t_2 denote the solutions, respectively, of the equations $\phi'(t)/\phi(t) = \epsilon$ and $\phi'(t)/\phi(t) = -\epsilon$. It follows from (2.8) that t_1 and t_2 exist and are uniquely determined for all sufficiently small ϵ. Let $a = \psi(t_1)$ and $b = \psi(-t_2)$, where ψ is defined by (2.6), with H replaced by ϵ therein.

The left side of (2.9) is equal to $P\left(\sum_1^n z_i \geqslant n\epsilon\right) + P\left(\sum_1^n (-z_i) \geqslant n\epsilon\right)$. Applications of Lemma 2.2 show that this equals $a_n^n + b_n^n$, where $a_n \to a$ and $b_n \to b$ as $n \to \infty$. It follows hence that $P(|S_n|/n \geqslant \epsilon) = c_n^n$, where $c_n \to c = \max\{a, b\}$. Consequently,

$$\lim_{n \to \infty} \{n^{-1} \log P(|N| > \epsilon/\lambda_n(\epsilon))\} = \log c, \qquad \dots \quad (2.13)$$

by (2.9). It follows from (2.9) and (2.13) by Lemma 2.3 that

$$\lim_{n \to \infty} \{n \lambda_n^2(\epsilon)\} = \frac{\epsilon^2}{-2 \log c}. \qquad \dots \quad (2.14)$$

It follows from the definition of t_1 and (2.8) that $t_1 = \epsilon/\sigma^2 + o(\epsilon)$ as $\epsilon \to 0$, where $\sigma^2 = \phi''(0) = E(z^2)$. Since $\phi(t) = 1 + t^2\sigma^2/2 + o(t^2)$ as $t \to 0$, we have $\log a = -\epsilon t_1 + \log \phi(t_1) = -\epsilon^2/2\sigma^2 + o(\epsilon^2)$. This last expansion is valid also for b, and therefore for c. The desired conclusion (2.10) now follows from (2.14). The equivalence of (2.10) with (2.11), (2.12) is a consequence of Lemma 2.3. This completes the proof.

3. Asymptotic efficiency of tests

In this section and the following one we consider the statistical framework described in the first paragraph of Section 1, but no parametric function g is specified. The following three conditions on the sample space X of a single observation x are assumed to hold.

Condition 3.1 : θ is identifiable, i.e. if θ_0 and θ_1 are points in Θ with $\theta_0 \neq \theta_1$ then $P(x \text{ in } A \mid \theta_0) \neq P(x \text{ in } A \mid \theta_1)$ for at least one measurable set $A \subset X$.

Condition 3.2 : The set of alternative distributions of x is dominated, i.e. there exists a σ-finite measure on X, say $\mu(A)$, and a non-negative function $f(x \mid \theta)$ on $X \times \Theta$ such that, for each θ in Θ, f is measurable in x, and such that

$$P(x \text{ in } A \mid \theta = \int_A f(x \mid \theta)\, d\mu \text{ for all } A \subset X. \quad \ldots \quad (3.1)$$

Define
$$z(x \mid \theta_1, \theta_0) = \log\, [f(x \mid \theta_1)/f(x \mid \theta_0)]. \quad \ldots \quad (3.2)$$

As in Section 1 in another context, 0/0 and ∞/∞ are here understood to be equal to 1.

Condition 3.3 : For any two points θ_0 and θ_1 in Θ,

$$E(z^2 \mid \theta_1) < \infty \quad \ldots \quad (3.3)$$

and
$$E(\exp\,(\varepsilon z) \mid \theta_1) < \infty \text{ for some } \varepsilon = \varepsilon(\theta_1, \theta_0) > 0 \quad \ldots \quad (3.4)$$

where z is given by (3.2).

Of these conditions, Condition 3.3 is the only one that is at all restrictive. The condition implies, in particular, that the set of alternative distributions of x is homogeneous, i.e. if $P(A \mid \theta) = 0$ for some θ then $P(A \mid \theta) = 0$ for all θ.

Let θ_0 and θ_1 be any two points of Θ with $\theta_0 \neq \theta_1$. In the remainder of this section we restrict attention to the problem of testing the simple hypothesis that $\theta = \theta_0$ against the simple alternative that $\theta = \theta_1$. For each n, let $x_{(n)}$ denote a sample $(x_1, x_2, \ldots, x_n)$ of n independent observations on x, and let $X_{(n)}$ denote the sample space of $x_{(n)}$, as in Section 1. A *test* is defined to be a sequence $\{W_n\} = \mathbf{W}$ say, such that W_n is a measurable subset of $X_{(n)}$. For any given test $\mathbf{W} = \{W_n\}$, let

$$\alpha_n(\mathbf{W}) \equiv \alpha(W_n) = P(x_{(n)} \text{ in } W_n \mid \theta_0),\ \beta_n(\mathbf{W}) \equiv \beta(W_n) = P(x_{(n)} \text{ not in } W_n \mid \theta_1). \quad \ldots \quad (3.5)$$

α_n and β_n are then the probabilities of errors of the first and second kinds, respectively, in using W_n as the critical region.

Definition 3.1 : $\mathcal{C}$ ($= \mathcal{C}(\theta_1)$) is the class of all tests W such that

$$\overline{\lim_{n\to\infty}} \{\beta_n(\mathrm{W})\} < 1. \qquad \ldots \quad (3.6)$$

For any test W in $\mathcal{C}$,

$$\underline{c}(\mathrm{W}) = \underline{\lim_{n\to\infty}} \{2n^{-1}\log[1/\alpha_n(\mathrm{W})]\}, \quad \bar{c}(\mathrm{W}) = \overline{\lim_{n\to\infty}} \{2n^{-1}\log[1/\alpha_n(\mathrm{W})]\}. \qquad \ldots \quad (3.7)$$

$\underline{c}[\bar{c}]$ is called the lower [upper] asymptotic slope of the test W.

The notion of asymptotic slope has been discussed by Bahadur (1960b) in the special case when the test $\mathrm{W} = \{W_n\}$ is based on a sequence of real valued statistics satisfying certain conditions. This special case is also discussed in the next section of this paper. For the present, let us note that the ratio of the slopes of two tests serves as their relative asymptotic efficiency in the following sense. Let $\mathrm{W}^{(1)} = \{W_n^{(1)}\}$ and $\mathrm{W}^{(2)} = \{W_n^{(2)}\}$ be two tests, and suppose for simplicity that $0 < \underline{c}(\mathrm{W}^{(i)}) = \bar{c}(\mathrm{W}^{(i)}) = c_i$ (say) $< \infty$, $i = 1, 2$. Given ϵ, $0 < \epsilon < 1$, let $M^{(i)}(\epsilon)$ be the least positive integer n such that $\alpha(W_n^{(i)}) \leqslant \epsilon$, where α is defined by (3.5). It then follows easily from the definition (3.7) of c_i that $M^{(i)}(\epsilon) = (-2 \log \epsilon)/(c_i + \delta_i)$, where $\delta_i \to 0$ as $\epsilon \to 0$. Consequently, $\lim_{n\to\infty} \{M^{(1)}(\epsilon)/M^{(2)}(\epsilon)\} = c_2/c_1$, i.e. c_2/c_1 is the limiting ratio of the sample sizes required by $\mathrm{W}^{(1)}$ and $\mathrm{W}^{(2)}$, respectively, to attain an arbitrarily small probability of an error of the first kind. In particular, $\mathrm{W}^{(1)}$ is more efficient (in the present sense) than $\mathrm{W}^{(2)}$ if $c_1 > c_2$.

The following Theorem 3.1 states in effect that, in the class $\mathcal{C}$, each one of a class of Neyman-Pearson likelihood ratio tests has the maximum slope, and that the numerical value of the maximum slope is $2H$, where

Definition 3.2 : $$H = H(\theta_1, \theta_0) = E[z(x\,|\,\theta_1, \theta_0)\,|\,\theta_1]. \qquad \ldots \quad (3.8)$$

It follows easily from Conditions 3.1—3.3, the definition (3.2) of z, and the inequality $\log t \leqslant t-1$ for $t \geqslant 0$, that we always have $0 < H < \infty$.

Let $\sigma^2 = \sigma^2(\theta_1, \theta_0)$ denote the variance of $z(x\,|\,\theta_1, \theta_0)$ when θ_1 obtains. It follows from Conditions 3.1 and 3.3 that $0 < \sigma^2 < \infty$. Let a be a constant, $-\infty < a < \infty$, and put $r_n = \exp[nH + \sqrt{n}\, a\, \sigma]$, where H is given by (3.8), $0 < r_n < \infty$. For each n, let $W_n^* = \{x_{(n)} : \prod_1^n f(x_i\,|\,\theta_1) \geqslant r_n \prod_1^n f(x_i\,|\,\theta_0)\}$, and let $\mathrm{W}^* = \{W_n^*\}$.

Theorem 3.1 : *(i)* W^* *is in* $\mathcal{C}$. *(ii)* $\underline{c}(\mathrm{W}^*) = \bar{c}(\mathrm{W}^*) = 2H$. *(iii)* $\bar{c}(\mathrm{W}) \leqslant 2H$ *for all* W *in* $\mathcal{C}$.

Proof : Write $z_i = z(x_i\,|\,\theta_1, \theta_0)$. Then by the definition of W_n^* and (3.5) we have $\beta(W_n^*) = P([\Sigma_1^n z_i - nH]/\sqrt{n}\,\sigma < a\,|\,\theta_1)$. Since H and σ^2 are the mean and variance of z when θ_1 obtains, it follows from the central limit theorem that

$$\lim_{n\to\infty} \beta(W_n^*) = P(N < a). \qquad \ldots \quad (3.9)$$

Since a is finite, it follows from (3.9) that W^* satisfies (3.6), and part (i) is established.

We have $\alpha_n(W_n^*) = P(\Sigma_1^n z_i \geqslant nH+\sqrt{n}\, a\,\sigma\,|\,\theta_0)$, by (3.5) and the definition of W_n^*. We shall apply Lemma 2.2 to this α_n. Putting $k_n = nH+\sqrt{n}\, a\,\sigma$, we see that (2.3) is satisfied. Since $E(e^{(1+t)z}|\theta_0) = E(e^{tz}|\theta_1)$ by (3.2), it follows from (3.4) that assumption (ii) preceding Lemma 2.1 is satisfied and that $t = 1$ is an interior point of the interval on which $\phi(t) = E(e^{tz}|\theta_0)$ is finite. Since $\phi'(t)/\phi(t) = E(ze^{tz}|\theta_0)/\phi(t)$, it follows from the present definitions of z and H that $\phi'(1)/\phi(1) = H$, so that assumption (iii) is also satisfied, with $t_0 = 1$ (cf.(2.4)). It is readily seen that assumption (iv) must also hold, otherwise Condition 3.1 would be violated. Thus Lemma 2.2 applies. In the present case, $\rho = \psi(t_0) = \exp[-t_0 H].\,\phi(t_0) = \exp[-H]$, so that $n^{-1}\log\alpha_n \to -H$ as $n\to\infty$. In view of (3.7), this establishes part (ii).

To prove part (iii), choose and fix a test W in $\mathcal{C}$. Then choose and fix constants a and $\epsilon > 0$ such that $\beta_n(\mathrm{W}) \leqslant P(N < a)-\epsilon$ for all sufficiently large n; the existence of a and ϵ is assured by (3.6). Let W* be defined as in the paragraph preceding the statement of Theorem 1. It then follows from (3.9) that $\beta(W_n) < \beta(W_n^*)$ for all sufficiently large n. Since W_n^* is a likelihood ratio test, it follows from the lemma of Neyman and Pearson that we must have $\alpha(W_n) > \alpha(W_n^*)$ for all sufficiently large n. Hence $\bar{c}(\mathrm{W}) \leqslant \bar{c}(\mathrm{W}^*)$ by (3.7). Since $\bar{c}(\mathrm{W}^*) = 2H$ by part (ii), this completes the proof of the theorem.

The quantity H has been studied by Kullback and Leibler (1951) in terms of the sample space of a single observation x. They showed, in particular, that if $y = S(x)$ is a statistic defined on X, and if $H_0(\theta_1, \theta_0)$ is the resulting quantity when x is replaced by y in the definition of H, then $H_0 \leqslant H$, with equality if and only if y is a sufficient statistic relative to the two-point parameter space $\{\theta_1, \theta_0\}$. Theorem 3.1 affords the following statistical interpretation of the stated theorem of Kullback and Leibler. Writing $y_i = S(x_i)$, suppose that instead of the original sequence $x_1, x_2, \ldots$ only the sequence $y_1, y_2, \ldots$ is made available to the statistician. Then the maximum asymptotic slope available (i.e. $2H_0$) is the same as before (i.e. $2H$) if S is a sufficient statistic, but is otherwise smaller.

4. The maximum slope of a standard sequence

In this section we consider the problem of testing the null hypothesis that $\theta = \theta_0$, where θ_0 is a given point of Θ, by means of suitable real valued statistics. Let $T_1, T_2, \ldots$ be a sequence such that T_n is a real valued measurable function of $x_{(n)}$. Write $P(T_n < u|\theta_0) = F_n(u)$.

Definition 4.1 : $\{T_n\}$ is said to be a standard sequence for testing $\theta = \theta_0$ if the following three conditions are satisfied. (I) $\lim_{n\to\infty} F_n(u) = F(u)$ for every u, where F is a continuous probability distribution function. (II) There exists a function on $(0, \infty)$ into $(0, \infty)$, say f, such that for any sequence $\{u_n\}$ with $u_n \to \infty$, $u_n^2/n \to d$, where $0 < d < \infty$, we have $2n^{-1}\log[1-F_n(u_n)] \to -f(d)$. (III) There exists a function b on Θ such that $b(\theta_0) = 0$ and $0 < b(\theta) < \infty$ for $\theta \neq \theta_0$, and such that $\{T_n/\sqrt{n}\}$

is a consistent estimate of b. If $\{T_n\}$ is a standard sequence, $f(b^2(\theta))$ is called its asymptotic slope when θ obtains, $(\theta \neq \theta_0)$.

Suppose that $\{T_n\}$ is a standard sequence. Then T_n has a limiting distribution if $\theta = \theta_0$, but $T_n \to \infty$ in probability if $\theta \neq \theta_0$, so that large values of T_n are significant when T_n is regarded as a testing statistic. Consequently, given $x_{(n)}$, $1-F_n(T_n(x_{(n)})) = L_n$ (say) is called the level attained by T_n in the given case, i.e. L_n is the probability of obtaining a value of T_n which is greater than or equal to the value actually observed. Write $K_n = -2 \log L_n$. It can be shown that in the null case K_n tends in distribution to a chi-square with two d.f., and that in the non-null case $K_n/n \to f(b^2(\theta))$ in probability.

This last result implies that the ratio of the asymptotic slopes of two alternative standard sequences is, roughly speaking, the inverse ratio of the sample sizes required in order to attain comparable levels of significance in the non-null case. A number of examples of this method of comparison of tests are given by Bahadur (1960a, 1960b). As is stated in these papers, although the method makes no explicit reference to the Neyman-Pearson theory of tests, there is a formal connection with the latter theory. This connection is exploited in the proof of Theorem 4.1 below.

A standard sequence $\{T_n\}$ is said to be non-degenerate at θ_1 if there exists a sequence of constants, $\{k_n\}$ say, such that

$$0 < \varliminf_{n\to\infty} P(T_n < k_n | \theta_1) \leqslant \varlimsup_{n\to\infty} P(T_n < k_n | \theta_1) < 1. \qquad \dots \quad (4.1)$$

Assuming Conditions 3.1, 3.2, and 3.3, we shall now prove

Theorem 4.1 : *If the standard sequence $\{T_n\}$ is non-degenerate at θ, its asymptotic slope cannot exceed $2H(\theta, \theta_0)$ when θ obtains.*

Proof : Choose and fix a $\theta_1 \neq \theta_0$, and suppose $\{T_n\}$ non-degenerate at θ_1. Then there exists a sequence $\{k_n\}$ such that (4.1) holds. For each n, let $W_n = \{x_{(n)} : T_n(x_{(n)}) \geqslant k_n\}$, and regard $\mathbf{W} = \{W_n\}$ as a test of θ_0 against θ_1 (cf. Section 3). Then $\mathbf{W}$ is in class $\mathcal{C}$, by (4.1) and Definition 3.1. We shall show that

$$\underline{c}(\mathbf{W}) = \bar{c}(\mathbf{W}) = f(b^2(\theta_1)). \qquad \dots \quad (4.2)$$

The desired conclusion will then follow from part (iii) of Theorem 3.1.

By condition (III) of Definition 4.1, $T_n/\sqrt{n} \to b(\theta_1)$ in probability, where $0 < b(\theta_1) < \infty$. It follows hence that

$$\lim_{n\to\infty} \{k_n/\sqrt{n}\} = b(\theta_1), \qquad \dots \quad (4.3)$$

for otherwise (4.1) would not hold. Now, $\alpha(W_n) = P(T_n \geqslant k_n | \theta_0) = 1-F_n(k_n)$. It follows, therefore, by (4.3) and condition (II) of Definition 4.1, that $2n^{-1} \log [1/\alpha(W_n)] = f(b^2(\theta_1))+o(1)$ as $n \to \infty$, i.e. that (4.2) holds. This completes the proof.

In general, there exists no standard sequence which is optimum in the sense that its slope equals $2H(\theta, \theta_0)$ for each θ in Θ. This is essentially because in general, for given n and given size α, there exists no uniformly most powerful critical region in $X_{(n)}$ for testing $\theta = \theta_0$ against the composite hypothesis $\theta \neq \theta_0$. A special case where such regions do exist, and consequently an optimum standard sequence also exists, is treated by Bahadur (1960a).

It seems fairly certain, however, that under general conditions there exists a standard or nearly standard sequence $\{T_n^*\}$, with slope c^* say, which is locally optimum in the sense that $c^*/2H$ tends to 1 as θ tends to θ_0 through non-null values, i.e. $\{T_n^*\}$ is nearly optimum throughout a neighbourhood of θ_0. This is argued in the subsequent paragraphs of this section. The notion of local optimality seems to be of considerable interest. One reason is that if θ is very distant from θ_0 it is implausible that a large sample will be drawn for testing purposes, so that asymptotic comparisons of alternative standard sequences (whatever such comparisons may be worth) might as well be confined to the immediate vicinity of θ_0. Consequently, a locally optimum sequence is, for practical purposes, an optimum sequence. This argument is, of course, parallel to the argument for letting ε tend to zero in the definition (1.3) of the relative efficiency of two estimates of a given parametric function g. Another reason is that (as is already suggested, and as will be clear from the following paragraphs) the problem of local optimality lies well within the relatively unexplored transition stage from the problem of efficient testing to that of efficient estimation.

Suppose that Θ is an open set in the k dimensional Euclidean space of points $\theta = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(k)})$, $1 \leqslant k < \infty$. Suppose that, for fixed $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}, \ldots, \theta_0^{(k)})$, $H(\theta, \theta_0)$ is a sufficiently smooth function of θ; to be specific, let D_i denote partial differentiation with respect to $\theta^{(i)}$, and assume that the second order derivative $D_r D_s [H(\theta, \theta_0)]$ exists and is a continuous function of θ, for all $r, s = 1, 2, \ldots, k$. Since $H(\theta_0, \theta_0) = 0$, and $H(\theta, \theta_0) > 0$ for $\theta \neq \theta_0$, it then follows from Taylor's theorem that with $I_{ij}(\theta_0) = \{D_i D_j H(\theta, \theta_0)\}_{\theta=\theta_0}$, and

$$Q(\theta, \theta_0) = \sum_{i,j=1}^{k} I_{ij}(\theta_0)\,(\theta^{(i)} - \theta_0^{(i)})\,(\theta^{(j)} - \theta_0^{(j)}), \qquad \ldots \quad (4.4)$$

we have

$$H(\theta, \theta_0) = \tfrac{1}{2}\, Q(\theta, \theta_0)\,[1 + \epsilon(\theta, \theta_0)] \qquad \ldots \quad (4.5)$$

where $\epsilon \to 0$ as $\theta \to \theta_0$. $\{I_{ij}(\theta_0)\}$ is necessarily a symmetric positive definite matrix.

It follows from the definition (3.8) of H that with $z = z(x|\theta, \theta_0)$ we have

$$H(\theta, \theta_0) = E(z\, e^z|\theta_0), \text{ and } E(e^z|\theta_0) = 1. \qquad \ldots \quad (4.6)$$

Assuming that $D_j[z]$ and $D_i D_j[z]$ exist for each x, and that the operators D_j and $D_i D_j$ commute with the expectation signs in (4.6), $(i, j = 1, 2, \ldots, k)$, it follows from the preceding definition of I_{ij} by a straightforward calculation that

$$I_{ij}(\theta_0) = E(u_i u_j \mid \theta), \text{ where } u_r = \{D_r[\log f(x \mid \theta)]\}_{\theta=\theta_0} \qquad \ldots \quad (4.7)$$

and $E(u_r \mid \theta_0) = 0$. Thus $[I_{ij}]$ coincides with the classical information matrix.

Now, for each n, let V_n be a measurable function on $X_{(n)}$ into k dimensional space, say $V_n(x_{(n)}) = (V_n^{(1)}(x_{(n)}), \ldots, V_n^{(k)}(x_{(n)}))$, such that, for each θ in Θ, $\sqrt{n}\,(V_n - \theta)$ tends in distribution to the k-variate normal distribution with mean zero and covariance matrix $\{I^{ij}(\theta)\} = \{I_{ij}(\theta)\}^{-1}$. It is known that these requirements are met if, for example, V_n = the m.l. estimate of θ based on $x_{(n)}$, and if certain regularity conditions (which need not be specified here) are satisfied. Let

$$T_n^*(x_{(n)}) = \sqrt{n \sum_{i,j=1}^{k} I_{ij}(\theta_0)(V_n^{(i)} - \theta_0^{(i)})(V_n^{(j)} - \theta_0^{(j)})}. \qquad \ldots \quad (4.8)$$

Then it is easy to see that $\{T_n^*\}$ satisfies condition (I) of Definition 4.1 with $F(u) = P(\chi_k < u)$, where χ_k^2 is a chi-square variable with k d.f., and that (III) is satisfied with $b(\theta) = \sqrt{Q}$, where Q is given by (4.4).

It seems very difficult to verify condition (II) in the general case, and to determine the function f^* associated with $\{T_n^*\}$ by that condition (assuming that such a function exists). Consequently, the exact level attained by T_n^* in a given case, L_n^* say, cannot be treated by the method described in the paragraph following Definition 4.1. However, given $x_{(n)}$, $P(\chi_k \geqslant T_n^*(x_{(n)})) = L_n^0$ say, is an approximation to L_n^* and this approximate level L_n^0 is easy to analyse. It can be shown that Lemma 2.3 remains valid if $|N|$ is replaced by χ_k (cf. Bahadur, 1960b). It follows hence that, with $K_n^0 = -2 \log L_n^0$, K_n^0/n tends in probability to Q. Thus, in a certain approximate sense, Q is the slope of $\{T_n^*\}$. It now follows from (4.5) that, again in a certain approximate sense, $\{T_n^*\}$ is a locally optimum sequence.

In order to show that $\{T_n^*\}$ is exactly a locally optimum sequence, it is necessary to show at least that $-2n^{-1} \log P(T_n \geqslant T_n^*(x_{(n)}) \mid \theta_0)$ tends in probability to $c^*(\theta)$ (say) when θ obtains, and that $c^*/Q \to 1$ as $\theta \to \theta_0$. It is to be hoped that this can be done by the methods used in the following section, provided $\{T_n^*\}$ is constructed according to (4.8) from the m.l. estimate of θ.

5. Theorems on estimation

In this section we consider the framework described in the first paragraph of Section 1, with g a given real valued function defined on Θ. It is assumed that Conditions 3.1, 3.2 and 3.3 are satisfied. In the statements of the additional conditions required in this section, θ_0 denotes an arbitrary but fixed point of Θ.

Condition 5.1: Given θ_0, the set

$$\{\theta : \theta \text{ in } \Theta,\ |g(\theta) - g(\theta_0)| = \epsilon\} = S_g(\epsilon, \theta_0) \text{ (say)} \qquad \ldots \quad (5.1)$$

is non-empty for all sufficiently small $\epsilon > 0$.

Definition 5.1: For any $\epsilon > 0$ such that $S_g(\epsilon, \theta_0)$ defined by (5.1) is non-empty,

$$J_g(\epsilon|\theta_0) = \inf\{H(\theta, \theta_0) : \theta \text{ in } S_g(\epsilon, \theta_0)\}, \quad \ldots \quad (5.2)$$

where H is given by (3.2) and (3.8). For any real $r > 0$,

$$K_g^{(r)}(\theta_0) = \varliminf_{\epsilon \to 0} \left\{\frac{J_g(\epsilon|\theta_0)}{\epsilon^r}\right\}. \quad \ldots \quad (5.3)$$

Although $K_g^{(r)}$ is always well defined, with $0 \leqslant K_g^{(r)} \leqslant \infty$, the interesting case is when

$$0 < K_g^{(r)}(\theta) < \infty. \quad \ldots \quad (5.4)$$

It is evident from (5.3) that, given θ, (5.4) can hold for at most one value of r. This useful value, if it exists, is usually $r = 2$, and occasionally $r = 1$. It should also be noted here that the useful value of r may vary with θ in a given estimation problem (cf. remark 2 in Section 6).

The following theorem, which gives bounds on the rate of convergence of any consistent estimate, is valid provided only that Conditions 3.1–3.3 and 5.1 are satisfied.

Theorem 5.1: *If* $\mathbf{T} = \{T_n\}$ *is a consistent estimate of* g *then, for every* θ *in* Θ *and every* $r > 0$,

$$\varliminf_{\epsilon \to 0} \varliminf_{n \to \infty} \{(n\,\epsilon^r)^{-1} \log P(|T_n - g(\theta)| \geqslant \epsilon|\theta)\} \geqslant -K_g^{(r)}(\theta). \quad \ldots \quad (5.5)$$

Proof : Choose and fix an r, and a θ_0 in Θ. We shall establish (5.5) at this θ_0. Choose $\epsilon > 0$ so small that $S_g(\epsilon, \theta_0)$ is non-empty, and let θ_1 be a point in the latter set. Let λ be a constant, $0 < \lambda < 1$, and put $\delta = \lambda\epsilon$.

For each n, let $W_n = \{x_{(n)} : |T_n - g(\theta_0)| \geqslant \delta\}$ and regard $\mathbf{W} = \{W_n\}$ as a test of θ_0 against θ_1 (cf. Section 3). Then

$$\beta(W_n) = P(|T_n - g(\theta_0)| < \delta|\theta_1) \quad \ldots \quad (5.6)$$

by (3.5). Now, $|g(\theta_1) - g(\theta_0)| = \epsilon$, so that

$$|T_n - g(\theta_0)| \geqslant |\epsilon - |T_n - g(\theta_1)|| \quad \ldots \quad (5.7)$$

for every $x_{(n)}$. Since $\delta = \lambda\epsilon < \epsilon$, it follows from (5.6), (5.7), and the consistency of T, that $\beta(W_n) \to 0$ as $n \to \infty$. Consequently, $\mathbf{W}$ is in the class $\mathcal{C}(\theta_1)$ of Definition 3.1. Hence $\bar{c}(\mathbf{W}) \leqslant 2H(\theta_1, \theta_0)$ by Theorem 3.1. It follows from the definitions of $\bar{c}$ and $\mathbf{W}$ that this last conclusion is

$$\varliminf_{n \to \infty} \{n^{-1} \log P(|T_n - g(\theta_0)| \geqslant \delta|\theta_0)\} \geqslant -H(\theta_1, \theta_0). \quad \ldots \quad (5.8)$$

Since θ_1 in $S_g(\epsilon, \theta_0)$ is arbitrary, it follows from (5.8) by (5.2) that

$$\underset{n\to\infty}{\underline{\lim}} \{(n\,\delta^r)^{-1} \log P(|T_n - g(\theta_0)| \geqslant \delta|\theta_0)\} \geqslant -\frac{J_g(\epsilon|\theta_0)}{\lambda^r \epsilon^r}. \qquad \ldots \quad (5.9)$$

Since ϵ is arbitrary, and $\delta = \lambda\epsilon \to 0$ continuously through positive values as $\epsilon \to 0$, it follows from (5.9) by (5.3) that at θ_0 the left side of (5.5) is not less than $-K_g^{(r)}(\theta_0)/\lambda^r$. If $K_g^{(r)}(\theta_0) = 0$ or ∞, this establishes (5.5) at θ_0; if $0 < K_g^{(r)}(\theta_0) < \infty$, the desired conclusion follows by letting the arbitrary λ tend to 1. This completes the proof.

It is easy to see from the preceding proof that (5.5) is valid not only for consistent estimates but for any $\{T_n\}$ such that $g(\theta_1) \neq g(\theta)$ implies $\overline{\lim}_{n\to\infty} P(|T_n - g(\theta)| \geqslant \delta|\theta_1) < 1$ for all positive $\delta < |g(\theta_1) - g(\theta)|$.

A different but equally effortless generalisation of Theorem 5.1 is to the case when g is not necessarily real valued. Suppose that g is a function on Θ into a metric space, and that Condition 5.1 is satisfied, with $|g(\theta_1) - g(\theta_0)|$ = the distance between $g(\theta_1)$ and $g(\theta_0)$. With the obvious definitions of a consistent estimate of g, and of $K_g^{(r)}(\theta)$, the preceding proof of (5.5) applies verbatim.

In the remainder of this section it will be necessary to assume that θ takes values in a Euclidean space, and that $H(\theta, \theta_0)$ and $g(\theta)$ are sufficiently smooth functions of θ. For simplicity, we shall consider only the case when θ is a real valued parameter. The parallel development when θ is a k-dimensional vector is discussed briefly in remark 4 of Section 6.

Condition 5.2: θ is real valued. Θ is an open interval. For θ in the neighbourhood of θ_0,

$$H(\theta, \theta_0) = \frac{(\theta-\theta_0)^2}{2} I(\theta_0) + o((\theta-\theta_0)^2), \qquad \ldots \quad (5.10)$$

where $0 < I(\theta_0) < \infty$. g is a differentiable function of θ.

It follows by taking $k = 1$ in the paragraphs containing (4.4)—(4.7) that, under certain regularity conditions, I defined by (5.10) coincides with the classical 'information concerning θ contained in x', but these additional conditions are not required at present.

Theorem 5.2: *If* $\mathrm{T} = \{T_n\}$ *is a consistent estimate of* g, *then* (1.5) *holds for each* θ *in* Θ.

Proof: Choose and fix a θ_0 in Θ. We shall establish the inequality (1.5) at θ_0. Since $0 < I(\theta_0) < \infty$, we may assume $g'(\theta_0) \neq 0$, for otherwise the inequality is trivial. Since $g'(\theta_0) \neq 0$, it is plain that Condition 5.1 is satisfied. Moreover, for each sufficiently small $\epsilon > 0$, we can choose a θ_ϵ in $S_g(\epsilon, \theta_0)$ so that $\theta_\epsilon \to \theta_0$ as $\epsilon \to 0$. Since $J_g(\epsilon|\theta_0) \leqslant H(\theta_\epsilon, \theta_0)$ for every ϵ by (5.2), we have

$$\frac{J_g(\epsilon|\theta_0)}{\epsilon^2} \leqslant \frac{H(\theta_\epsilon, \theta_0)}{[g(\theta_\epsilon)-g(\theta_0)]^2}. \qquad \ldots \quad (5.11)$$

It follows from (5.10) by letting $\epsilon \to 0$ in (5.11) that

$$K_0^{(2)}(\theta_0) \leqslant \tfrac{1}{2} I(\theta_0) \,/\, [g'(\theta_0)]^2. \qquad \dots \quad (5.12)$$

It follows readily from (1.4), (5.5) with $r = 2$, and (5.12) that (1.5) holds at θ_0, and this completes the proof.

Theorem 5.2 states in effect that the asymptotic effective variance (Definition 1.3) of a consistent estimate of g cannot be less than $[g']^2/nI$. The remainder of this section is devoted to showing that, under certain additional conditions, the asymptotic effective variance of the m.l. estimate is equal to this lower bound. The conclusion to be established is, in a sense, stronger and more precise than the conclusion that the m.l. estimate of g is a c.a.n. estimate with asymptotic variance $[g']^2/nI$. The regularity conditions which we shall require are essentially a combination of Wald's conditions for the consistency of the m.l. estimate of θ (Wald, 1949), of Cramér's conditions for the asymptotic normality of a root of the likelihood equation (Cramér, 1946) and of conditions required in order to apply the lemmas of Section 2 in certain ways.

Let $L(\theta|x) = \log f(x|\theta)$, where f is given by (3.1), and let $D \equiv \partial/\partial\theta$.

Condition 5.3 : (i) For each x in X, $L(\theta|x)$ is a thrice differentiable function of θ. (ii) With

$$u(x|\theta_0) = \{DL\}_{\theta=\theta_0} \text{ and } v(x|\theta_0) = \{D^2L\}_{\theta=\theta_0}, \qquad \dots \quad (5.13)$$

we have

$$E(u|\theta_0) = 0,\ E(u^2|\theta_0) = -E(v|\theta_0) = I(\theta_0). \qquad \dots \quad (5.14)$$

(iii) There exists a measurable function $w(x|\theta_0)$, and a neighbourhood of θ_0, such that

$$|D^3L| \leqslant w(x|\theta_0) \qquad \dots \quad (5.15)$$

for every x in X and every θ in the neighbourhood. (iv) The m.g.fs of u, v, and w exist when θ_0 obtains.

In Condition 5.3 (iv) and in Condition 5.4 below, the finite existence of m.g.fs is required only in a neighbourhood of the origin.

In the following, for any real valued function h defined on Θ and any set $\Omega \subset \Theta$, we write $h(\Omega) = \sup\{h(\theta) : \theta \text{ in } \Omega\}$. In particular, with $z(x|\theta, \theta_0)$ defined by (3.2),

$$z(x|\Omega, \theta_0) = \sup\{z(x|\theta, \theta_0) : \theta \text{ in } \Omega\}. \qquad \dots \quad (5.16)$$

Since $z(x|\theta, \theta_0)$ is measurable in x for each θ and, according to Condition 5.3(i), continuous in θ for each x, it is easy to see that $z(x|\Omega, \theta_0)$ is measurable (but not necessarily finite valued).

Condition 5.4 : The m.g.f. of $z(x|\Theta, \theta_0)$ exists when θ_0 obtains.

Since $0 \leqslant z(x|\Theta, \theta_0) \leqslant \infty$, this condition is equivalent to $E(e^{tz} \mid \theta_0) < \infty$ for some $t > 0$. The reader may verify that if X is a discrete sample space, and $\sum_x [P(\{x\}|\theta_0)]^p < \infty$ for some power $p < 1$, then Condition 5.4 is satisfied.

In order to state the next and final regularity condition, suppose $\Theta = \{\theta : a < \theta < b\}$, where $-\infty \leqslant a < b \leqslant \infty$.

Condition 5.5 : For each x in X, $f(x|\theta)$ tends to a limit as θ tends to a or to b. With

$$f(x|a) = \lim_{\theta \to a} \{f(x|\theta)\}, \quad f(x|b) = \lim_{\theta \to b} \{f(x|\theta)\} \qquad \dots \quad (5.17)$$

we have $\mu\{x : f(x|a) \neq f(x|\theta_0)\} > 0$ and $\mu\{x : f(x|b) \neq f(x|\theta_0)\} > 0$.

Now, for any $n = 1, 2, \dots$ and any $x_{(n)} = (x_1, x_2, \dots, x_n)$ in $X_{(n)}$, let $L_n(\theta|x_{(n)})$ denote the logarithm of the likelihood function given $x_{(n)}$, i.e.

$$L_n(\theta|x_{(n)}) = \log \prod_{i=1}^{n} f(x_i|\theta) = L(\theta|x_1) + \dots + L(\theta|x_n). \qquad \dots \quad (5.18)$$

Let $M_n(x_{(n)})$ be the (possibly empty) set of points θ in Θ such that $L_n(\theta|x_{(n)}) = L_n(\Theta|x_{(n)})$.

Definition 5.2: A sequence $\hat{\mathbf{T}} = \{\hat{T}_n\}$ is said to be an m.l. estimate of θ if, for each n, $\hat{T}_n$ is a measurable function on $X_{(n)}$ into Θ, and if $\hat{T}_n(x_{(n)})$ is in $M_n(x_{(n)})$ whenever the latter set is non-empty. A sequence $\hat{\mathbf{U}} = \{\hat{U}_n\}$ is said to be an m.l. estimate of $g(\theta)$ if $\hat{U}_n \equiv g(\hat{T}_n)$, where $\hat{T}_n$ is an m.l. estimate of θ.

We can now state

Lemma 5.1: *There exists an m.l. estimate of* θ.

Given n and $x_{(n)}$, we can surely choose a point $\hat{T}_n$ in Θ, in such a way that $\hat{T}_n$ is in M_n if the latter set is non-empty. Lemma 5.1 asserts that this choice can be exercised (for those $x_{(n)}$ for which $M_n(x_{(n)})$ is empty or consists of more than one point) so that the resulting function of $x_{(n)}$ is measurable. The proof consists in giving a constructive definition of $\hat{T}_n$ based on the continuity of the likelihood function and the separability of Θ. This proof is omitted because it is rather long and uninteresting, and because the arguments of the following paragraphs apply to any given m.l. estimate. (cf. LeCam, 1956).

In what follows, $\{\hat{T}_n\}$ is a given m.l. estimate of θ. It is assumed that θ_0 obtains, where θ_0 is an arbitrary but fixed point of Θ.

Lemma 5.2: *Given* $h > 0$, *there exists* ρ, $0 < \rho < 1$, *such that*

$$P(|\hat{T}_n - \theta_0| \geqslant h) < \rho^n \qquad \dots \quad (5.19)$$

for all sufficiently large n.

This lemma implies the strong consistency of the m.l. estimate of θ. Its proof is along the lines of Wald's proof of consistency (Wald, 1949).

Proof : Consider a θ in Θ with $\theta \neq \theta_0$. We shall show that

$$E(e^{tz(x|\theta,\theta_0)}) < 1 \quad \text{for } 0 < t < 1. \qquad \dots \quad (5.20)$$

Letting $\phi(t)$ denote the expectation in (5.20), we have $\phi(0) = \phi(1)$, by the definition of z and the present assumption that θ_0 obtains. Since ϕ is an m.g.f., ϕ is convex in [0, 1]. Suppose that $\phi(t_1) = 1$ for some t_1 in (0, 1). It then follows from convexity that $\phi = 1$ for all t in (0, 1) so that $\phi''(t) = E(z^2 e^{tz}) = 0$ in that interval. Hence $P(z = 0) = 1$, i.e. $P(f(x|\theta) = f(x|\theta_0)|\theta_0) = 1$. Consequently, $\mu\{f(x|\theta) \neq f(x|\theta_0)\} = 0$, and this is contrary to Condition 3.1.

Next, let $f(x|a)$ be defined by (5.17) and $z(x|a, \theta_0)$ as usual by (3.2), $-\infty \leqslant z \leqslant \infty$. We shall show that (5.20) continues to hold with θ replaced by a. Let $\theta_1, \theta_2, \dots$ be a sequence in Θ which tends to a. Writing ϕ_i for ϕ when θ is replaced by θ_i in the preceding paragraph, we have $\phi_i(1) = 1$ for each i. Writing $z = z(x|a, \theta_0)$, it follows hence by (5.17) and Fatou's lemma that with $\phi(t) \equiv E(e^{tz})$ we have $\phi(1) \leqslant 1$. Let $\chi = 1$ if $z > -\infty$ and $\chi = 0$ otherwise, and let $\phi_0(t) = E(\chi e^{tz})$. Then $\phi_0(t) \leqslant \phi(t)$ for all t, with equality for $t > 0$. Suppose that there exists a t_1 with $0 < t_1 < 1$ such that $\phi(t_1) \geqslant 1$. We then have $\phi_0(t_1) \geqslant 1$, $\phi_0(0) \leqslant 1$, and $\phi_0(1) \leqslant 1$. Since ϕ_0 is convex on [0, 1], it follows that $\phi_0 = 1$ on this interval. Now $\phi_0(0) = 1$ implies $E(\chi) = 1$, i.e. $P(z > -\infty) = 1$. Consequently, $\phi(t) = \phi_0(t) = 1$ on [0, 1]. It follows hence, as in the preceding paragraph, that $P(z = 0) = 1$, and hence $\mu\{f(x|a) \neq f(x|\theta_0)\} = 0$, which contradicts the second part of Condition 5.5. The argument of this paragraph obviously remains valid when a is replaced by b.

It is therefore established that (5.20) holds for each $\theta \neq \theta_0$ in $\Theta^* = \{\theta : a \leqslant \theta \leqslant b\}$, with $z(x|a, \theta_0)$ and $z(x|b, \theta_0)$ defined as in the preceding paragraph. Now choose and fix a t such that $0 < t < 1$ and such that $E(\exp[tz(x|\Theta, \theta_0)]) < \infty$; such a t exists by Condition 5.4. It then follows from (5.20), by an application of Lebesgue's dominated convergence theorem, that for each θ in Θ^* with $\theta \neq \theta_0$ there exists an interval containing θ, $\mathcal{J}(\theta)$ say, such that with

$$z^*(x|\theta) = \sup \{z(x|\theta_1, \theta_0) : \theta_1 \text{ in } \mathcal{J}(\theta)\}, \quad \rho(\theta) = E(e^{tz^*}) \qquad \dots \quad (5.21)$$

we have $\rho(\theta) < 1$, and such that $\mathcal{J}(\theta)$ is open in Θ^*.

Given $h > 0$, let Ω be the set $\{\theta : \theta \text{ in } \Theta^*, |\theta - \theta_0| \geqslant h\}$. If Ω is empty, (or indeed if $\Omega - \{a\} - \{b\}$ is empty) then $|\hat{T}_n - \theta_0| \geqslant h$ is an impossible event, so that (5.19) holds for every n with $\rho = \frac{1}{2}$ (say). Suppose now that Ω is non-empty. Since Ω is a compact subset of Θ^*, and Ω does not contain θ_0, it is possible to find a finite number of points in Ω, say $\theta_1, \theta_2, \dots, \theta_k$ such that $\Omega \subset \mathcal{J}(\theta_1) + \mathcal{J}(\theta_2) + \dots + \mathcal{J}(\theta_k)$, where $\mathcal{J}(\theta)$ is defined for each $\theta \neq \theta_0$ in the preceding paragraph.

Now consider a fixed n and $x_{(n)}$, and suppose that $|\hat{T}_n - \theta_0| \geqslant h$. Suppose first that there do exist θ values in Θ for which $L_n(\theta) = L_n(\Theta)$, i.e. M_n is non-empty. Then $\hat{T}_n$ is in M_n and also in Ω. Consequently, $L_n(\Omega) \geqslant L_n(\hat{T}_n) = L_n(\Theta) \geqslant L_n(\theta_0)$.

Thus

$$L_n(\Omega|x_{(n)}) \geqslant L_n(\theta_0|x_{(n)}). \qquad \dots \quad (5.22)$$

Suppose next that M_n is empty, i.e. $L_n(\hat{T}_n) \neq L_n(\Theta)$. In that case, $L_n(\Theta) = \max\{L_n(a), L_n(b)\} > L_n(\theta_0)$; hence (5.22) continues to hold, since a and b are included in Ω. Thus $|\hat{T}_n - \theta_0| \geqslant h$ always implies (5.22).

Since $\Omega \subset \mathcal{I}(\theta_1)+\dots+\mathcal{I}(\theta_k)$, (5.22) implies

$$\max_{1 \leqslant j \leqslant k} \{L_n(\mathcal{I}(\theta_j)|x_{(n)})\} \geqslant L_n(\theta_0|x_{(n)}). \qquad \dots \quad (5.23)$$

Now, for any θ, $L_n(\mathcal{I}(\theta)|x_{(n)}) - L_n(\theta_0|x_{(n)})$ is less than or equal to $\sum_{i=1}^{n} z^*(x_i|\theta)$, by (5.18) and (5.21). Consequently, (5.23) implies

$$\max_{1 \leqslant j \leqslant k} \left\{\sum_{i=1}^{n} z^*(x_i|\theta_j)\right\} \geqslant 0. \qquad \dots \quad (5.24)$$

Let $A_n^{(j)}$ denote the event that $\sum_{i=1}^{n} z^*(x_i|\theta_j) \geqslant 0$. Then (5.24) is equivalent to $A_n^{(1)} + A_n^{(2)} + \dots + A_n^{(k)}$. Consequently, by the preceding paragraphs,

$$P(|\hat{T}_n - \theta_0| \geqslant h) \leqslant \sum_{j=1}^{k} P(A_n^{(j)}). \qquad \dots \quad (5.25)$$

It follows from the definition of $A_n^{(j)}$ by taking $H = 0$ and $z = z^*(x|\theta_j)$ in Lemma 2.1 that $P(A_n^{(j)}) \leqslant [\rho(\theta_j)]^n$, where $\rho(\theta_j)$ is given by (5.21). Let $\rho_0 = \max\{\rho(\theta_1), \dots, \rho(\theta_k)\}$. Then $\rho_0 < 1$, and the right side of (5.25) does not exceed $k\,\rho_0^n$. Choose a ρ such that $\rho_0 < \rho < 1$. Then $k\rho_0^n < \rho^n$ for all sufficiently large n. Consequently, by (5.25), (5.19) holds for all sufficiently large n. This establishes Lemma 5.2.

It might appear at first sight that a refinement of the preceding argument would lead to the best possible ρ in (5.19) and that by determining the dependence of this ρ on h it would be possible to determine the asymptotic effective variance of $\hat{T}_n$, but this is not the case. The reason is essentially that the proof of Lemma 5.2 reduces Θ in effect to a finite set. When Θ is finite, the m.l. estimate of θ has necessarily a discrete distribution, and statistics with discrete distributions are liable to have large effective standard deviations. It is therefore necessary to take into account the fact that Θ is a continuum. This is done in the proof of the following lemma by using Condition 5.3 and exhibiting the m.l. estimate as a root of the likelihood equation.

Lemma 5.3: *Given δ, $0 < \delta < I(\theta_0)$, there exists a $\rho < 1$ such that, for every $\epsilon > 0$,*

$$P(|\hat{T}_n - \theta_0| \geqslant \epsilon) \leqslant P(|\zeta_n| \geqslant \epsilon[I(\theta_0) - \delta]) + \rho^n \qquad \ldots \quad (5.26)$$

for all sufficiently large n, where

$$\zeta_n = \frac{\sum_{i=1}^{n} u(x_i|\theta_0)}{}. \qquad \ldots \quad (5.27)$$

Proof : Write

$$\eta_n = I(\theta_0) + \frac{\sum_{i=1}^{n} v(x_i|\theta_0)}{n}, \quad \xi_n = \frac{\sum_{i=1}^{n} w(x_i|\theta_0)}{n} \qquad \ldots \quad (5.28)$$

where v and w are given by (5.13) and (5.15). Choose and fix an $h > 0$ such that Θ contains the interval $\{\theta : |\theta - \theta_0| < h\}$, (5.15) holds for every θ in this last interval, and $\delta/h > E(w|\theta_0)$. Since v and w possess m.g.fs when θ_0 obtains (Condition 5.3), and $E(v) = -I(\theta_0)$ by (5.14), it follows from (5.28) by means of Lemma 2.1 that

$$P(\eta_n \geqslant \frac{\delta}{2}) \leqslant \rho_1^n, \; P(\eta_n \leqslant -\frac{\delta}{2}) \leqslant \rho_2^n, \; P(\xi_n > \frac{\delta}{h}) \leqslant \rho_3^n \qquad \ldots \quad (5.29)$$

for every n, where each ρ_i is less than 1.

Suppose that for given n and $x_{(n)}$ we have

$$|\hat{T}_n - \theta_0| < h, \; L_n(\hat{T}_n) = L_n(\Theta) \, . \qquad \ldots \quad (5.30)$$

Write $\Delta_n(\theta) = D[L_n(\theta)]$. Since $\hat{T}_n$ is in Θ, and Θ is open, it follows from (5.30) by Taylor's theorem that there exists a θ^* with $|\theta^* - \theta_0| < |\hat{T}_n - \theta_0| < h$ such that $\Delta_n(\theta_0) + (\hat{T}_n - \theta_0)\,\Delta_n'(\theta_0) + \frac{1}{2}(\hat{T}_n - \theta_0)^2\,\Delta_n''(\theta^*) = \Delta_n(\hat{T}_n) = 0$. It follows hence by referring to (5.13), the definition of Δ_n, and (5.27), (5.28), that

$$(\hat{T}_n - \theta_0)[I(\theta_0) + r_n] = \zeta_n \qquad \ldots \quad (5.31)$$

where

$$|r_n| \leqslant |\eta_n| + \tfrac{1}{2} h \, \xi_n. \qquad \ldots \quad (5.32)$$

It is not required here that r_n be a measurable function of $x_{(n)}$, or that (5.30) be a measurable event in $X_{(n)}$.

Let A_n denote the event $|\hat{T}_n - \theta_0| \geqslant h$, B_n the event $L_n(\hat{T}_n) \neq L_n(\Theta)$, and C_n the event $|\eta_n| + \frac{1}{2} h \, \xi_n \geqslant \delta$. As is shown in the proof of Lemma 5.2, each of the events A_n and B_n implies (5.22), and the probability of (5.22) is $\leqslant \rho_0^n$ for all sufficiently large n, where $\rho_0 < 1$. C_n implies at least one of the three events whose probabilities are considered in (5.29). We conclude that there exists a $\rho < 1$ and a measurable event E_n such that $A_n + B_n + C_n$ implies E_n, and such that $P(E_n) \leqslant \rho^n$ for all sufficiently large n.

For any given ϵ, the left side of (5.26) does not exceed $P(x_{(n)}$ *not in* $E_n, |\hat{T}_n - \theta_0| \geqslant \epsilon) + P(E_n)$. It follows hence from (5.31), (5.32), and the preceding paragraph that (5.26) holds for all sufficiently large n, and this completes the proof.

We can now establish

Theorem 5.3: *If* $\{\hat{U}_n\}$ *is an m.l. estimate of* g, *then* (1.6) *holds whenever* $g'(\theta) \neq 0$.

Proof : Choose and fix a δ, $0 < \delta < I(\theta_0)$, and write $a = I(\theta_0) - \delta$. It follows from (5.27) by Lemma 2.4 and Condition 5.3 that, given $\rho < 1$, $n^{-1} \log P(|\zeta_n| \geqslant \epsilon a)$ tends to a limit $> \log \rho$ as $n \to \infty$, provided only that $\epsilon > 0$ is sufficiently small. It follows hence from (5.26) that, with $A_{n,\epsilon} = \{x_{(n)} : |\hat{T}_n - \theta_0| \geqslant \epsilon\}$, we have

$$\overline{\lim_{n\to\infty}} \{n^{-1} \log P(A_{n,\epsilon})\} \leqslant \overline{\lim_{n\to\infty}} \{n^{-1} \log P(|\zeta_n| \geqslant \epsilon a)\} \qquad \ldots \quad (5.33)$$

for all sufficiently small ϵ. Multiply (5.33) with $1/\epsilon^2$, and let $\epsilon \to 0$. It then follows from (5.14) and (5.27) by Lemma 2.4 that

$$\overline{\lim_{\epsilon\to 0}} \quad \overline{\lim_{n\to\infty}} \{(n\,\epsilon^2)^{-1} \log P(A_{n,\epsilon})\} \leqslant -\tfrac{1}{2} \frac{a^2}{I(\theta_0)}. \qquad \ldots \quad (5.34)$$

Since $a = I(\theta_0) - \delta$, and δ is arbitrary, it follows from (5.34) that

$$\overline{\lim_{\epsilon\to 0}} \quad \overline{\lim_{n\to\infty}} \{(n\,\epsilon^2)^{-1} \log P(A_{n,\epsilon})\} \leqslant -\tfrac{1}{2} I(\theta_0). \qquad \ldots \quad (5.35)$$

In view of (1.4), (5.35) is equivalent to (1.6) at θ_0 in the special case $g(\theta) \equiv \theta$. To treat the general case, suppose $\{\hat{U}_n\}$ is an m.l. estimate of g. Then $\hat{U}_n \equiv g(\hat{T}_n)$, where $\{\hat{T}_n\}$ is an m.l. estimate of θ. Suppose that $g(\theta_0) \neq 0$. Choose and fix a $\lambda > 1$. It is easy to see that, for all sufficiently small $\epsilon > 0$, $|g(\theta) - g(\theta_0)| \geqslant \epsilon$ implies $|\theta - \theta_0| \geqslant \delta$, where $\delta = \epsilon / \lambda |g'(\theta_0)|$. Let $B_{n,\epsilon} = \{x_{(n)} : |\hat{U}_n - g(\theta_0)| \geqslant \epsilon\}$. Then $B_{n,\epsilon}$ implies $A_{n\delta}$ so that $P(B_{n,\epsilon}) \leqslant P(A_{n,\delta})$. Since $\delta/\epsilon \equiv 1/\lambda |g'|$, it follows from (5.35) that

$$\overline{\lim_{\epsilon\to 0}} \quad \overline{\lim_{n\to\infty}} \{(n\,\epsilon^2)^{-1} \log P(B_{n,\epsilon})\} \leqslant -\tfrac{1}{2} \frac{I(\theta_0)}{\lambda^2 [g'(\theta_0)]^2}. \qquad \ldots \quad (5.36)$$

Since $\lambda > 1$ is arbitrary, (5.36) implies

$$\overline{\lim_{\epsilon\to 0}} \quad \overline{\lim_{n\to\infty}} \{(n\,\epsilon^2)^{-1} \log P(B_{n,\epsilon})\} \leqslant -\tfrac{1}{2} \frac{I(\theta_0)}{[g'(\theta_0)]^2}. \qquad \ldots \quad (5.37)$$

It follows from (5.37), the definition of $B_{n,\epsilon}$, and (1.4) that (1.6) holds at θ_0, and this completes the proof of Theorem 5.3.

6. Further remarks on estimation

The following remarks concern the definitions and conclusions of Sections 1 and 5.

1. It is clear from (1.4), and from the proofs of Section 5, that the present study of estimation is based in effect on the logarithm of the probability of a deviation of ϵ or more from the value being estimated. However, when this last probability is the criterion, the interposition of the logarithm means that only the dominant features of the estimates being compared are taken into account. In particular, there may exist two alternative estimates $\mathbf{T} = \{T_n\}$ and $\mathbf{U} = \{U_n\}$, such that both $\mathbf{T}$ and $\mathbf{U}$ are efficient estimates of g according to Definition 1.2, but such that $\mathbf{T}$ is actually more efficient in the sense that

$$\overline{\lim_{\epsilon \to 0}} \ \overline{\lim_{n \to \infty}} \left\{ \frac{P(|T_n - g(\theta)| \geqslant \epsilon | \theta)}{P(|U_n - g(\theta)| \geqslant \epsilon | \theta)} \right\} < 1 \quad \ldots \quad (6.1)$$

for all θ. This is the case if, for example, x is real valued and normally distributed with mean θ and variance 1, $g(\theta) = \theta$, $T_n = (x_1 + \ldots + x_n)/n = \bar{x}_n$, and $U_n = \bar{x}_{m_n}$, where $m_1, m_2, \ldots$ is a sequence of positive integers such that $m_n \leqslant n$ for every n, $\lim_{n \to \infty} \{m_n/n\} = 1$, and $\lim_{n \to \infty} \{n - m_n\} = \infty$.

It is possible to formulate asymptotic relative efficiencies in terms of limits such as the left side of (6.1), but the analysis then required seems very difficult, even in the simplest cases.

2. The following example shows that, in general, the r for which (5.4) holds depends on θ.

Suppose that x is real valued, $X = \{x : 0 < x < 1\}$, $\Theta = \{\theta : 0 < \theta < 1\}$, μ = Lebesgue measure on X (cf. Condition 3.2), and

$$f(x|\theta) = \begin{cases} \dfrac{1}{2\theta} & \text{if } \ 0 < x \leqslant \theta \\ \\ \dfrac{1}{2(1-\theta)} & \text{if } \ \theta < x < 1. \end{cases} \quad \ldots \quad (6.2)$$

Let $g(\theta) \equiv \theta$. It is readily seen that Conditions 3.1, 3.2, 3.3 and 5.1 are satisfied, so that Theorem 5.1 applies.

Let us write K instead of K_g, since here g is the identity function. A straightforward but lengthy calculation, which is omitted, shows that

$$K^{(2)}(\theta) = \begin{cases} 2 & \text{if } \ \theta = \frac{1}{2} \\ \\ \infty & \text{if } \ \theta \neq \frac{1}{2}, \end{cases} \quad \ldots \quad (6.3)$$

and that

$$K^{(1)}(\theta) = \begin{cases} 0 & \text{if } \theta = \frac{1}{2} \\ \frac{1}{2\delta}\left[\frac{(2\delta-1)}{(1-\delta)} + \log\left(\frac{1-\delta}{\delta}\right)\right] & \text{if } \theta \neq \frac{1}{2}, \end{cases} \quad \ldots \quad (6.4)$$

where $\delta = \max\{\theta, 1-\theta\}$.

It is plain from (6.2) that θ is the median of the distribution of x. Consequently, if $T_n^* =$ the median of the sample values $\{x_1, x_2, \ldots, x_n\}$, $\mathbf{T}^* = \{T_n^*\}$ is a consistent estimate of θ. By expressing the distribution of T_n^* in terms of the binomial distribution, and applying Lemma 2.2 to the latter distribution, it can be shown that

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \{(n\epsilon^2)^{-1} \log P(|T_n^* - \theta| \geqslant \epsilon | \theta)\} = -\frac{1}{2\delta^2} \quad \ldots \quad (6.5)$$

for every θ. It follows from (6.3) and (6.5) that $\mathbf{T}^*$ is efficient when $\theta = \frac{1}{2}$. It does *not* follow from (6.4) and (6.5) alone that $\mathbf{T}^*$ is inefficient for other values of θ. This is because Theorem 5.1 only provides a bound which may or may not be attainable in a given case.

It would be interesting to know whether, in the present case, there exists an estimate which is efficient according to Definition 1.2, and if so, whether efficiency can be established by means of Theorem 5.1.

3. Let $\mathbf{T} = \{T_n\}$ be a c.a.n. estimate of g, and suppose that, when θ obtains, the asymptotic variance of T_n is $v(\theta)/n$, where $0 < v < \infty$. It was generally believed, following the work of R. A. Fisher, that we must then have $v(\theta) \geqslant [g'(\theta)]^2/I(\theta)$ for all θ. This belief was shown to be erroneous by J. L. Hodges, who constructed examples of estimates which were superefficient in the sense that $v(\theta) \leqslant [g'(\theta)]^2/I(\theta)$ for all θ, and

$$v(\theta) < [g'(\theta)]^2/I(\theta) \quad \ldots \quad (6.6)$$

for some values of θ. A detailed study of superefficiency is given by LeCam (1953), who showed that in general the set of values θ for which (6.6) holds must be a set of Lebesgue measure zero.

It follows from Theorem 5.2 that the phenomenon of superefficiency depends on a certain lack of uniformity in the approach to normality. To see this, let λ be a positive constant. Then

$$\lim_{n \to \infty} P(\sqrt{n}|T_n - g(\theta)| \geqslant \lambda | \theta) = P(|N| \geqslant \lambda/\sqrt{v(\theta)}) \quad \ldots \quad (6.7)$$

by the asymptotic normality of $\mathbf{T}$. It follows from (6.7) and Definition 1.1 that

$$\lim_{n \to \infty} \left\{ n\, \tau_\theta^2(T_n, \frac{\lambda}{\sqrt{n}}, \theta) \right\} = v(\theta) \quad \ldots \quad (6.8)$$

for every λ and θ. Suppose, for a given θ, that (6.8) holds uniformly in λ. It then follows from (1.5) that (6.6) cannot hold for that θ.

4. Suppose that Θ is an open set of the k dimensional Euclidean space of points $\theta = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(k)})$, and that the conditions stated in the paragraph

containing (4.4)—(4.5) are satisfied. Suppose that the given real valued function $g(\theta)$ possesses continuous partial derivatives $\partial g/\partial\theta^{(i)} = h_i(\theta)$ (say) for $i = 1, 2, \ldots, k$.

The conditions just stated generalise Condition 5.2. If Conditions 3.1—3.3 are also satisfied,

$$\underline{\lim}_{\epsilon\to 0}\ \underline{\lim}_{n\to\infty}\ \{n\,\tau_g^2(T_n, \epsilon, \theta)^2\} \geqslant \sum_{i,j=1}^{k} I^{ij}(\theta)\, h_i(\theta)\, h_j(\theta) \quad \ldots \quad (6.9)$$

for any consistent estimate of g. The proof of (6.9) is exactly the same as that of Theorem 5.2, with θ_ϵ in the proof of Theorem 5.2 so chosen that the vector $\theta_\epsilon - \theta_0$ is proportional to the vector $(r_1, r_2, \ldots, r_k)$ which minimises $\sum_{i,j=1}^{k} I_{ij}(\theta_0)\, r_i r_j$ subject to the condition $\sum_{i=1}^{k} h_i(\theta_0)\, r_i = 1$ (say).

It is a little more difficult, but quite possible, to generalise Theorem 5.3, i.e., to show (under certain conditions) that if $\{\hat{U}_n\}$ is an m.l. estimate of g then

$$\overline{\lim}_{\epsilon\to 0}\ \overline{\lim}_{n\to\infty}\ \{n\,\tau_g^2(\hat{U}_n, \epsilon, \theta)\} \leqslant \sum_{i,j=1}^{k} I^{ij}(\theta)\, h_i(\theta)\, h_j(\theta) \quad \ldots \quad (6.10)$$

whenever the right side of (6.10) is positive. The main difficulty is in formulating a satisfactory generalisation of Condition 5.5; once this is done the proof of (6.10) proceeds along the same lines as that of Theorem 5.3.

REFERENCES

BAHADUR, R. R. (1960a): Simultaneous comparison of the optimum and sign tests of a normal mean. *Hotelling Festschrift*, Stanford University Press, (Stanford).

——— (1960b): Stochastic comparison of tests. *Ann. Math. Stat.*, **31** pp.—.

BAHADUR, R. R. and RANGA RAO, R. (1960): On deviations of the sample mean. *Ann. Math. Stat.*, **31**, pp.—.

BARANKIN, E. W. and GURLAND, J. (1951): On asymptotically normal efficient estimates. University of California Publications in Statistics. Vol. 1, No. 6, 89–130.

BASU, D. (1956): On the concept of asymptotic efficiency. *Sankhyā*, **17**, 193.

CHERNOFF, H. (1952): A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.*, **23**, 493–507.

CRAMÉR, H. (1946): *Mathematical Methods of Statistics*, Princeton University Press.

FELLER, W. (1957): *An Introduction to Probability Theory and its Applications*, Vol. 1, 2nd ed., John Wiley and Sons, (New York).

FISHER, R. A. (1922): On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, London, Series A, **222**, 309–368.

——— (1925): Theory of statistical estimation. *Proc. Cambridge Phil. Soc.*, **22**, 700–725.

KALLIANPUR, G. B. and RAO, C. R. (1955): On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā*, **15**, 331.

KULLBACK, S. and LEIBLER, R. A. (1951): On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86.

LECAM, L. (1953): On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. University of California Publications in Statistics. Vol. 1, No. 11, 277–330.

——— (1956): On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

NEYMAN, J. (1949): Contribution to the theory of the χ^2 test. Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, 239–273.

SAVAGE, L. J. (1954): *The Foundations of Statistics*, John Wiley and Sons, (New York).

WALD, A. (1949): Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, **20**, 595-601.

Paper received : February, 1960.