DISTRIBUTION OPTIMALITY AND SECOND-ORDER EFFICIENCY*

OF TEST PROCEDURES

R. R. Bahadur

University of Chicago

and

J. C. Gupta

Indian Statistical Institute

It has been shown, under certain conditions, by Bahadur, Chandra, and Lambert (1982) that in the non-null case the best possible asymptotic distribution for the level attained by a test statistic is a certain lognormal distribution, and that the level of the likelihood ratio statistic has this optimal asymptotic distribution. We describe a technical generalization of this theory; in the present generalization the best possible asymptotic distribution of the standardized log-level is that of the maximum of a family of normally distributed variables. It is pointed out that these considerations yield a corresponding generalization concerning the asymptotic expansion of the log-size of the best critical region when the power against a given alternative is a specified constant.

1. Introduction.

In the following sections S is a sample space of points s, and A is a σ-field of subsets of S. Θ is a parameter space of points θ and, for each

*This research was supported in part by National Science Foundation Grant No. MCS 8301459.

AMS 1980 subject classifications: 60F10, 62F05, 62G20.

Key words and phrases: significance testing, asymptotic distribution of p-values, second-order efficiency, likelihood ratio statistics, large deviations.

θ in Θ, P_θ is a probability on $\mathbf{A}$. n is an index taking values in the set of positive integers and, for each n, $\mathbf{B}_n$ is a subfield of A. We may think of n as the sample size or cost of the experiment $\{(S,\mathbf{B}_n,P_\theta): \theta \text{ in } \Theta\}$ concerning θ, but no particular relation is required of the fields $\mathbf{B}_n$ corresponding to different values of n. Θ_o is a subset of Θ, and the null hypothesis under test is that some θ in Θ_0 obtains. $\Theta_1 = \Theta - \Theta_0$ denotes the non-null set of parameter points. We assume that for each n, δ in Θ_1, and θ in Θ_0, the probability P_δ is dominated by P_θ on $\mathbf{B}_n$, i.e.,

(1) $$dP_\delta = R_n(s:\delta,\theta)\ dP_\theta \text{ on } \mathbf{B}_n$$

where R_n is $\mathbf{B}_n$ measurable and $0 \leq R_n < \infty$. Let

(2) $$K_n(s:\delta,\theta) = n^{-1} \log R_n(s:\delta,\theta),\ -\infty \leq K_n < \infty \ .$$

We also assume that there exists a constant $K(\delta,\theta)$ such that, as $n \longrightarrow \infty$,

(3) $$K_n(s:\delta,\theta) \longrightarrow K(\delta,\theta) \text{ in } P_\delta\text{-probability}.$$

Then, necessarily, $0 \leq K \leq \infty$.

The present framework is a version of the frameworks in Bahadur and Raghavachari (1972); Bahadur, Gupta, and Zabell (1980); and Bahadur, Chandra, and Lambert (1982). As may be seen from discussions and illustrative examples in these papers, which are henceforth referred to as [BR], [BGZ], and [BCL], and in references therein, K is a generalized Kullback-Leibler information number and it plays a central role in large deviation theories of testing and estimation.

2. Asymptotic distribution of the level attained.

For each n let T_n be an extended real valued B_n measurable function defined on S; T_n is to be thought of as a test statistic, large values of T_n being significant. Let

(4) $$G_n(t) = \sup\{P_\theta(T_n \geq t): \theta \text{ in } \Theta_0\}$$

for $-\infty \leq t \leq \infty$, and

(5) $$L_n(s) = G_n(T_n(s)), \quad 0 \leq L_n \leq 1 .$$

Then L_n is the level attained by T_n in testing Θ_0. It is known that in the null case L_n is uniform or superuniform over [0,1], i.e.

(6) $$P_\theta(L_n \leq \alpha) \leq \alpha \quad \text{for } 0 \leq \alpha \leq 1, \quad \theta \text{ in } \Theta_0 .$$

To consider the distribution of L_n in the non-null case, choose and fix δ in Θ_1.

With K defined by (1), (2), and (3), let

(7) $$\nu(\delta) = \inf\{K(\delta,\theta): \theta \text{ in } \Theta_0\} .$$

Assumption 1. $0 \leq \nu(\delta) < \infty$, and the set

(8) $$\Gamma_\delta = \{\theta: \theta \text{ in } \Theta_0, K(\delta,\theta) = \nu(\delta)\}$$

is non-empty.

For each point γ in Γ_δ, let

(9) $$Z_n(s:\delta,\gamma) = n^{1/2}[\nu(\delta) - K_n(s:\delta,\gamma)] \; .$$

<u>Assumption 2.</u> For each finite set $\{\gamma_1,\ldots,\gamma_m\}$ in Γ_δ, the distribution of $(Z_n(s:\delta,\gamma_1),\ldots,Z_n(s:\delta,\gamma_m))$ under P_δ converges to a (possibly singular) normal distribution centered at the origin in m-dimensional Euclidean space.

It is assumed in [BCL] that the point γ which minimizes $K(\delta,\theta)$ over Θ_0 is unique, i.e. Γ_δ is a singleton. In this case Assumptions 1 and 2 become the main assumptions of [BCL] and Assumptions 3 and 4 below are satisfied trivially. It should be added that non-uniqueness seems to be the exception rather than the rule. We think, however, that the present generalization is of interest because it provides insights into some of the difficulties and complications involved in general studies of tests of composite null hypotheses; see also Section 3.

For γ_1, γ_2 in Γ_δ, let $C_\delta(\gamma_1,\gamma_2)$ denote the asymptotic covariance of $Z_n(s:\delta,\gamma_1)$ and $Z_n(s:\delta,\gamma_2)$. There exists a measurable space Ω of points ω, a probability P on the measurable sets of Ω, and for each γ in Γ_δ a real-valued random variable $Y_\delta(\omega:\gamma)$ such that $\{Y_\delta(\omega:\gamma): \gamma \in \Gamma_\delta\}$ is a mean-zero Gaussian process with covariance function C_δ. Let

(10) $$V_\delta(\omega) = \sup\{Y_\delta(\omega:\gamma): \gamma \text{ in } \Gamma_\delta\} \; ,$$

$-\infty < V_\delta < \infty$. It is not necessary to specify Ω, P, and the $\{Y_\delta\}$, but we require the existence of a version of these entities such that the following assumption holds; see Fernique (1974) and references therein for general sufficient conditions on C_δ.

<u>Assumption 3.</u> Γ_δ is a separable metric space. With probability one, the sample function $Y_\delta(\omega:\gamma)$ is continuous on Γ_δ, and $V_\delta(\omega) < \infty$.

Now choose a statistic T_n for each n, and with L_n defined by (4) and (5), let

(11) $$M_n(s:\delta) = [\log L_n(s) + n\nu(\delta)]/n^{1/2}\,.$$

THEOREM 1. For each z, $-\infty < z < \infty$,

(12) $$\limsup_{n \to \infty} P_\delta(M_n(s:\delta) \leq z) \leq P(V_\delta \leq z).$$

Proof. First consider a fixed n. Let k and α be positive constants and γ a point in Γ_δ. It follows from (1) and (6) with $\theta = \gamma$, exactly as in the proof on page 6 of [BCL], that with $R_n = R_n(s\colon \delta,\gamma)$ we have $P_\delta(L_n \leq \alpha,\ R_n \leq k) \leq k\alpha$. It follows hence that if $\gamma_1,\ldots,\gamma_m$ are points in Γ_δ and $R_n^{(m)} = \min\{R_n(s:\delta,\gamma_i)\colon 1 \leq i \leq m\}$, then $P_\delta(L_n \leq \alpha,\ R_n^{(m)} \leq k) \leq mk\alpha$. Hence $P_\delta(L_n \leq \alpha) \leq P_\delta(R_n^{(m)} > k) + mk\alpha$. By letting $k = (n\alpha)^{-1}$ and $\alpha = \exp(-n\nu + n^{1/2}z)$ in this last inequality, it follows from (2), (9), and (11) that

(13) $$P_\delta(M_n \leq z) \leq P_\delta(Z_n^{(m)} < z + a_n) + (m/n)$$

where $Z_n^{(m)} = \max\{Z_n(s:\delta,\gamma_i)\colon\ 1 \leq i \leq m\}$ and $a_n = (\log n)/n^{1/2}$.

Since the maximum co-ordinate of the point in R^m is a continuous function of the point, it follows from Assumption 2 that $Z_n^{(m)}$ converges in distribution to $Y_\delta^{(m)} = \max\{Y_\delta(\omega:\gamma_i)\colon\ 1 \leq i \leq m\}$ as $n \to \infty$. It follows hence from (13) that, for any $\epsilon > 0$,

(14) $$\limsup_{n \to \infty} P_\delta(M_n \leq z) \leq P(Y_\delta^{(m)} \leq z + \epsilon).$$

The inequality (14) holds with any choice of $\gamma_1, \gamma_2,\ldots,\gamma_m$. Now let $\{\gamma_1, \gamma_2,\ldots\}$ be a dense subset of Γ_δ. It then follows from Assumption 3 and the definition of $Y_\delta^{(m)}$ that $Y_\delta^{(m)} \to V_\delta$ pointwise and therefore in distribution as $m \to \infty$; consequently $\limsup_m P(Y_\delta^{(m)} \leq z + \epsilon) \leq P(V_\delta \leq z + 2\epsilon)$. Since ε is arbitrary, it follows hence from (14) that (12) holds for all z.

Remark 1. It is plain from the proof that the bound (12) is uniform in T_n, i.e., with $Q_n(\delta:z)$ the supremum of $P_\delta(M_n \leq z)$ over all $\mathbf{B}_n$ measurable statistics T_n, $\limsup_n Q_n(\delta:z) \leq P(V_\delta \leq z)$ for all z.

A particular choice of statistic T_n for each n is said to be optimal in the sense of weak exact slopes (in short, w-optimal) against δ if $n^{-1}\log L_n(s) \rightarrow -\nu(\delta)$ in P_δ-probability; T_n is optimal in the sense of asymptotic distributions (in short, d-optimal) against δ if $M_n(s:\delta) \rightarrow V_\delta$ in distribution under P_δ. Since $P(-\infty < V_\delta < \infty) = 1$ by assumption, it is clear from (11) that d-optimality implies w-optimality. The following theorem shows that the converse holds in the class of statistics such that $\log L_n$ has an asymptotic normal distribution with mean and variance proportional to n.

THEOREM 2. Suppose that under P_δ

$$\log L_n(s) \text{ is } AN(-n\mu(\delta),\ n\sigma^2(\delta)) \tag{15}$$

as $n \rightarrow \infty$, where $0 \leq \mu(\delta) < \infty$, $0 \leq \sigma^2(\delta) < \infty$. Then either $\mu(\delta) < \nu(\delta)$, or $\mu(\delta) = \nu(\delta)$ and V_δ is an $N(0,\sigma^2(\delta))$ variable.

Proof. It follows from (15) that $n^{-1}\log L_n \rightarrow -\mu$ in probability. Hence, by [BR] or by arguments in [BCL], $\mu \leq \nu$. Suppose that $\mu = \nu$. Then, by (11) and (15), $M_n \rightarrow N(0,\sigma^2)$ in distribution. It follows hence from Theorem 1 that

$$P(N(0,\sigma^2) \leq z) \leq P(V_\delta \leq z) \tag{16}$$

for all z. Let γ be a point in Γ_δ. Then $Y_\delta(\omega:\gamma) \leq V_\delta(\omega)$ by (10), so $P(N(0,\sigma^2) \leq z) \leq P(Y_\delta(\gamma) \leq z)$ for all z, by (16). Since $Y_\delta(\gamma)$ is normal with mean 0 and variance $C_\delta(\gamma,\gamma)$, we must have $C_\delta(\gamma,\gamma) = \sigma^2$. Thus $Y_\delta(\gamma)$ is an $N(0,\sigma^2)$ variable. By (10) we now have $P(V_\delta \leq z) \leq P(Y_\delta(\gamma) \leq z)$ $= P(N(0,\sigma^2) \leq z)$ for all z; it follows hence from (16) that V_δ is $N(0,\sigma^2)$.

Remark 2. The conclusion that V_δ is a normal variable is equivalent to $P(Y_\delta(\gamma_1) = Y_\delta(\gamma_2)) = 1$ for all γ_1, γ_2 in Γ_δ; the conclusion therefore implies that the uniqueness assumption of [BCL] is essentially satisfied.

Remark 3. Suppose $\nu(\delta) = 0$. It then follows from Theorem 2 by letting $T_n \equiv 0$ (say) for each n that $P(V_\delta = 0) = 1$. It follows hence from Theorem 1 and $\nu = 0$ that, for any choice of T_n, not only $n^{-1}\log L_n$ but $n^{-1/2}\log L_n$ converges to 0 in P_δ-probability. The case $\nu = 0$ is therefore rather hopeless and is not considered further.

Remark 4. It is easy to construct examples, even with Γ_δ a singleton, of statistics T_n which are w-optimal but not d-optimal against δ. Of course, in such examples, (15) does not hold. Cf. Remarks 10 in Section 3.

We now describe sufficient conditions in order that a particular statistic T_n be d-optimal against a given δ. Let

(17) $$U_{n,\delta}(s) = n^{1/2}[\nu(\delta) - T_n(s)].$$

Condition 1. $U_{n,\delta}$ is asymptotically stochastically smaller than V_δ when δ obtains, i.e.

(18) $$\liminf_{n\to\infty} P_\delta(U_{n,\delta} < z) \geq P(V_\delta < z) \text{ for } -\infty < z < \infty .$$

Condition 2. With G_n defined by (4) there exists a function g_n, $-\infty \leq g_n < \infty$, such that

(19) $$n^{-1}\log G_n(t) \leq -t + g_n(t) \text{ for } -\infty < t < \infty$$

and such that with

(20) $$h_n(k:\delta) = n^{1/2}\sup\{g_n(t): |t - \nu(\delta)| < k/n^{1/2}\}$$

we have

(21) $$\limsup_{n \to \infty} h_n(k:\delta) \leq 0 \text{ for each } k,\ 0 < k < \infty .$$

THEOREM 3. If δ obtains, and T_n satisfies Conditions 1 and 2, then

(22) $$U_{n,\delta} \to V_\delta \text{ in distribution,}$$

(23) $$\log L_n = -n\nu(\delta) + n^{1/2}U_{n,\delta} + o(n^{1/2}) \text{ in probability,}$$

and T_n is d-optimal.

The proof of Theorem 3 is along the lines of the proof of Proposition 2.8 in [BCL] and is omitted. We note here that (22) holds for any sequence of extended random variables $U_{n,\delta}$ if and only if (18) and

(24) $$\limsup_{n \to \infty} P_\delta(U_{n,\delta} < z) \leq P(V_\delta < z) \text{ for } -\infty < z < \infty$$

are both satisfied. We note also from (1) and (6) that $-\infty < \log L_n \leq 0$ with P_δ-probability one, so the $o(n^{1/2})$ term in (23) is well-defined for each n even if $|U_{n,\delta}| = \infty$ for some sample points.

In order to apply Theorem 3 to likelihood ratio and related statistics, it is convenient to introduce here a rather natural additional assumption concerning the framework itself. With Z_n defined by (1), (2) and (9) let

(25) $$V_{n,\delta}(s) = \sup\{Z_n(s:\delta,\gamma): \gamma \text{ in } \Gamma_\delta\}$$

and suppose that $V_{n,\delta}$ is B_n measurable, $-\infty \leq V_{n,\delta} \leq \infty$. It is plain from Assumptions 2 and 3 that $V_{n,\delta}$ is asymptotically stochastically larger than V_δ when δ obtains, i.e. (24) holds with each $U_{n,\delta}$ replaced by $V_{n,\delta}$.

Assumption 4. $V_{n,\delta} \rightarrow V_\delta$ in distribution when δ obtains.

This assumption holds if, for example, Γ_δ is compact metric, each $Z_n(s:\delta,.)$ is a random element in the space of real-valued continuous functions on Γ_δ, this last space is equipped with the topology of uniform convergence, and the distribution of the element $Z_n(s:\delta,.)$ under P_δ converges to the distribution of the element $Y_\delta(\omega:.)$; see, e.g., Billingsley (1968).

In the following Remarks 5-7 we consider Conditions 1 and 2 of Theorem 3 for three likelihood ratio (LR) statistics $T^*_{n,\delta}$, $T^0_{n,\delta}$, and $\hat{T}_n$; the corresponding variables $U_{n,\delta}$ defined by (17) are denoted by $U^*_{n,\delta}$, $U^0_{n,\delta}$, and $\hat{U}_{n,\delta}$ respectively.

Remarks 5. Suppose for the moment that the null set is Γ_δ and δ is the singleton alternative. Then

$$T^*_{n,\delta}(s) = \inf\{K_n(s:\delta,\gamma): \gamma \text{ in } \Gamma_\delta\} \tag{26}$$

is a version of the relevant LR statistic. It is plain from (17), (25) and (26) that $U^*_{n,\delta}$ is $V_{n,\delta}$, so Condition 1 for $T^*_{n,\delta}$ is equivalent to Assumption 4. Condition 2 is always satisfied by $T^*_{n,\delta}$ when Γ_δ is the null set. To see this, we note first that, for any θ in Θ_0, (1) implies $P_\theta(R_n(\delta:\theta) \geq k) \leq k^{-1}$ for $0 < k < \infty$; hence for γ in Γ_δ, $P_\gamma(T^*_{n,\delta} \geq t) \leq P_\gamma(K_n(\delta,\gamma) \geq t) \leq \exp(-nt)$ for each real t, so Condition 2 holds with $g_n(t) \equiv 0$. Thus Assumption 4 is sufficient for d-optimality.

Remarks 6. For testing the given Θ_0 against a singleton δ

(27) $$T^0_{n,\delta}(s) = \inf[K_n(s:\delta,\theta): \theta \text{ in } \Theta_0\}$$

is the LR statistic corresponding to $T^*_{n,\delta}$. It follows from the argument in Remark 5 that, with Θ_0 as the null set, $T^0_{n,\delta}$ always satisfies Condition 2 with $g_n(t) \equiv 0$; consequently, Condition 1 is sufficient for d-optimality. It is plain from (26) and (27) that $U^0_{n,\delta} \geq U^*_{n,\delta} = V_{n,\delta}$. It follows hence from Stigler's Proposition 2.5 in [BCL] that Condition 1 holds for $T^0_{n,\delta}$ if and only if Assumption 4 holds and $U^0_{n,\delta} - U^*_{n,\delta} \longrightarrow 0$ in P_δ-probability, i.e.

(28) $$n^{1/2}[T^*_{n,\delta} - T^0_{n,\delta}] \longrightarrow 0 \text{ in } P_\delta\text{-probability.}$$

In (28) the indeterminate differences $\infty - \infty$ and $(-\infty) - (-\infty)$ are understood to be 0. Needless to say, if $T^0_{n,\delta}$ is known explicitly, it may be a simple matter to verify directly that $U^0_{n,\delta} \longrightarrow V_\delta$ in distribution.

Remarks 7. For testing the given Θ_0 against every δ in Θ_1 the LR statistic is

(29) $$\hat{T}_n(s) = \sup\{T^0_{n,\delta}(s):\delta \text{ in } \Theta_1\}.$$

For any particular non-null δ, $T^0_{n,\delta} \leq \hat{T}_n$ and hence $\hat{U}_{n,\delta} \leq U^0_{n,\delta}$. It follows that if $T^0_{n,\delta}$ satisfies Condition 1 at δ then so does $\hat{T}_n$. Verification of Condition 2 for $\hat{T}_n$ is, perhaps, the most challenging of the verifications under discussion, but the condition is usually found to hold in regular cases (cf. examples in [BR] and [BCL]); the underlying reason is that Condition 2 is satisfied by $\hat{T}_n$ whenever Θ_1 is finite, and discretization is operative in regular cases with infinite Θ_1.

The following is a simple example where Γ_δ is not a singleton and V_δ is not a normal variable. Suppose that $s = (x_1,x_2,\ldots)$ is a sequence of i.i.d. random vectors in the plane R^2, with each x_i normally distributed with mean vector θ and covariance matrix the identity. Let Θ be the plane, and

suppose Θ_0 is the unit circle $\{\theta:||\theta|| = 1\}$. For each n let $(S, \mathcal{B}_n)$ represent the sample space of $(x_1,\ldots,x_n)$. Then $K(\delta,\theta) = ||\delta - \theta||^2/2$.
If $\delta \neq (0,0)$, Γ_δ is a singleton and the considerations of [BCL] apply. Suppose $\delta = (0,0)$; then Assumption 1 holds with $\nu(\delta) = 1/2$ and $\Gamma_\delta = \Theta_0$. It is easily seen that, with $\langle\cdot,\cdot\rangle$ the Euclidean inner product, $Z_n(s:\delta,\gamma)$ equals $\langle\gamma, n^{1/2}\bar{x}_n\rangle$ for $\gamma \in \Theta_0$. Let Ω be the plane of points $\omega = (\omega_1,\omega_2)$ with ω_1 and ω_2 independent N(0,1) variables, and let $Y_\delta(\omega:\gamma) = \langle\gamma,\omega\rangle$. Then, for each n, the distribution of $\{Z_n(\delta:\gamma): \gamma \in \Theta_0\}$ coincides with that of $\{Y_\delta(\gamma): \gamma \in \Theta_0\}$ so Assumption 2 holds; since $\langle\gamma,\omega\rangle$ is continuous in γ and $V_\delta(\omega) = ||\omega||$, Assumption 3 holds, and V_δ is a χ_2 variable. Here $V_{n,\delta} = ||n^{1/2}\bar{x}_n||$ is a χ_2 variable for each n when δ obtains, so Assumption 4 also holds. Since $\Gamma_\delta = \Theta_0$, $T^*_{n,\delta} \equiv T^0_{n,\delta}$ and it follows from Remarks 6 and 7 that $\hat{T}_n$ satisfies Condition 1 of Theorem 3. To verify Condition 2 for $\hat{T}_n$, let $W_{n,\theta}$ be the LR statistic for testing the singleton θ against all alternatives. Then $W_{n,\theta}$ is a $\chi^2_2/2n$ variable when θ obtains.
Now, $\theta \in \Theta_0$ implies $\hat{T}_n \leq W_{n,\theta}$, and hence $P_\theta(\hat{T}_n \geq t) \leq P_\theta(W_{n,\theta} \geq t) = \exp(-nt)$ for $0 < t < \infty$; it follows hence that $\hat{T}_n$ satisfies Condition 2 with $g_n \equiv 0$. In the present example, $\hat{T}_n = (||\bar{x}_n|| - 1)^2/2$, and it is possible to verify Conditions 1 and 2 directly for any non-null δ.

Remark 8. In the preceding Remarks 5-7 and Example, and in [BCL], Conditions 1 and 2 are used as convenient sufficient conditions; in fact, in a certain sense they are also necessary for d-optimality. Corresponding to any statistic T_n there exists a statistic $\tilde{T}_n$ which is a non-decreasing function of T_n, with $\tilde{L}_n \equiv L_n$, such that T_n is d-optimal against a given δ if and only if $\tilde{T}_n$ satisfies Conditions 1 and 2 at δ; $\tilde{T}_n = n^{-1}\log L_n^{-1}$ is such a statistic.

3. Asymptotic expansion of the log-size.

The preceding section concerns descriptive significance testing; the statistician chooses a test statistic T_n and computes and records or even reports the level L_n attained by T_n. In this section we consider some related aspects of behaviorist testing; the statistician chooses a critical function ϕ_n and rejects the hypothesis, or accepts it, with respective probabilities ϕ_n and $1 - \phi_n$.

In the framework of Section 1, for each n let $\phi_n(s)$ be a $\mathbf{B}_n$ measurable function such that $0 \leq \phi_n \leq 1$. The size of ϕ_n in testing Θ_0, say $\alpha(\phi_n)$, is (see, e.g., Lehmann (1959))

$$\alpha(\phi_n) = \sup\{E_\theta(\phi_n): \theta \text{ in } \Theta_0\} . \tag{30}$$

Choose and fix a non-null δ and a constant β, $0 < \beta < 1$. We require ϕ_n to have power at least β against δ, i.e.

$$E_\delta(\phi_n) \geq \beta . \tag{31}$$

It has been known for some time that, in typical cases, (31) with equality implies $\log \alpha(\phi_n) = -n\mu + o(n)$ for some $\mu > 0$; more recently, several authors have obtained various refinements of this first-order expansion in various examples and contexts; see Section 1 of [BCL] and references given there. In this section we obtain the second-order expansion for the best critical function in the general case, and show that the log-size of a critical function based on a d-optimal statistic has this expansion.

It is assumed henceforth that Assumptions 1-3 of Section 2 hold with $\nu(\delta) > 0$ and $P(V_\delta = 0) \neq 1$. It follows from Assumption 3 by the remarkable results in Tsirel'son (1975) that the β-quantile of V_δ is uniquely determined, i.e.,

$$\begin{aligned} \sup\{z:\ P(V_\delta \leq z) < \beta\} &= \inf\{z:\ P(V_\delta < z) > \beta\} \\ &= q(\delta:\beta) \text{ say,} \end{aligned} \tag{32}$$

$-\infty < q < \infty$, and that q is a continuous function of β for $0 < \beta < 1$.

Let $v_\delta = \inf\{z: P(V_\delta > z) > 0\}$ and $p_\delta = P(V_\delta = v_\delta)$. If $v_\delta = -\infty$ or, more generally, if $p_\delta = 0$, then q is strictly increasing on (0,1) with range (v_δ, ∞); but if $p_\delta > 0$ then $p_\delta < 1$, $q = v_\delta$ for $0 < \beta \leq p_\delta$, and q is strictly increasing on $(p_\delta, 1)$ with range (v_δ, ∞).

Let $\xi_n(\delta:\beta)$ be the infimum of $\alpha(\phi_n)$ over all $\mathbf{B}_n$ measurable ϕ_n such that (31) holds. This infimum is generally attained, but not necessarily by a critical function based on the LR statistics $\hat{T}_n$ or even $T^0_{n,\delta}$, since the Neyman-Pearson lemma does not extend to tests of composite hypotheses.

COROLLARY 1. As $n \to \infty$,

$$\log \xi_n(\delta:\beta) \geq -n\nu(\delta) + n^{1/2} q(\delta:\beta) + o(n^{1/2}). \tag{33}$$

Proof. Let u be a random variable uniformly distributed over the interval $[0,1] = I$ say, independent of s, and let $S^* = I \times S$ be the space of points $s^* = (u,s)$. With $\mathbf{B}$ the Borel field in I let $\mathbf{A}^* = \mathbf{B} \times \mathbf{A}$, and let P^*_θ be the probability measure on $\mathbf{A}^*$ when θ in Θ obtains. For each n let $\mathbf{B}^*_n = \mathbf{B} \times \mathbf{B}_n$, and call S^*, $\mathbf{A}^*$, $\{\mathbf{B}^*_n\}$, and $\{P^*_\theta\}$ the augmented framework. It is easy to see that all our assumptions, including Assumptions 1-3, continue to hold in the augmented framework, with $\nu(\delta)$ retaining its original value and V_δ its original distribution.

For a particular n let ϕ_n be a $\mathbf{B}_n$ measurable critical function such that (31) holds and

$$\alpha(\phi_n) < 1 . \tag{34}$$

Since $\xi_n \leq \beta < 1$, the additional restriction (34) involves no loss of generality. Let

$$T_n^*(s^*) = \begin{cases} 1 & \text{if} \quad u \le \phi_n(s) \\ 0 & \text{otherwise}. \end{cases} \tag{35}$$

Then

$$P_\theta^*(T_n^* = 1) = E_\theta(\phi_n) \text{ for all } \theta \text{ in } \Theta . \tag{36}$$

With $L_n^*(s^*)$ the level attained by T_n^* in testing Θ_0, it follows from (30), (35), and (36) that

$$L_n^* = \begin{cases} \alpha(\phi_n) & \text{if} \quad T_n^* = 1 \\ 1 & \text{if} \quad T_n^* = 0 . \end{cases} \tag{37}$$

It is plain from (34) and (37) that $L_n^* \le \alpha(\phi_n)$ if and only if $T_n^* = 1$. Hence

$$P_\delta^*(L_n^* \le \alpha(\phi_n)) = E_\delta(\phi_n) \ge \beta \tag{38}$$

by (31) and (36). We refer the reader to Kallenberg (1981, 1983) for other uses of the augmented framework in relating behaviorist and descriptive theories of testing.

Now choose a z such that $P(V_\delta \le z) < \beta$, and let

$$\lambda_n = \exp[-n\nu + n^{1/2} z] . \tag{39}$$

It then follows from Theorem 1 and Remark 1 in Section 2 that $\limsup_n P_\delta^*(L_n^* \le \lambda_n) < \beta$ uniformly in $\mathbf{B}_n^*$- measurable statistics T_n^* . In particular, there exists m such that $n > m$ implies $P_\delta^*(L_n^* \le \lambda_n) < \beta$ for all ϕ_n under consideration; hence $\alpha(\phi_n) > \lambda_n$ for all ϕ_n, by (38). Hence $n > m$ implies $\xi_n \ge \lambda_n$, i.e. $\log \xi_n \ge -n\nu + n^{1/2} z$. Hence $\liminf_n n^{-1/2} [\log \xi_n + n\nu] \ge z$. Since z is arbitrary, it follows from (32) that (33) holds.

Next, let T_n be a $\mathbf{B}_n$ measurable statistic defined on S, and let L_n be the level attained by T_n in testing Θ_0. Let $a_n = a_n(\delta,\beta)$ be the constant

$$P_\delta(L_n < a_n) \leq \beta \leq P_\delta(L_n \leq a_n) \ , \quad 0 \leq a_n \leq 1 \ , \tag{40}$$

and let

$$\phi_n(s) = \begin{cases} 1 & \text{if} \quad L_n \leq a_n \\ 0 & \text{if} \quad L_n > a_n \end{cases} . \tag{41}$$

Then ϕ_n satisfies condition (31). Of course, this ϕ_n based on T_n depends on δ and β, even if T_n itself does not.

COROLLARY 2. Suppose that there exists a T_n which is d-optimal in testing Θ_0 against δ. Then

$$\log \xi_n(\delta:\beta) = -n\nu(\delta) + n^{1/2} q(\delta:\beta) + o(n^{1/2}). \tag{42}$$

Moreover, if ϕ_n is based on T_n according to (40) and (41) then ϕ_n is efficient to second-order, i.e., $\log \alpha(\phi_n) = \log \xi_n + o(n^{1/2})$.

Proof. Choose a d-optimal T_n and let ϕ_n be based on T_n as above. Since $\alpha(\phi_n) \geq \xi_n$ for each n, and since (33) holds, it will suffice to show that

$$\log \alpha(\phi_n) \leq - n\nu(\delta) + n^{1/2} q(\delta:\beta) + o(n^{1/2}). \tag{43}$$

It follows from (41) and (6) that, for θ in Θ_0, $E_\theta(\phi_n) = P_\theta(L_n \leq a_n) \leq a_n$; hence $\alpha(\phi_n) \leq a_n$ for each n, by (30).

Choose a z such that $P(V_\delta < z) > \beta$, and let λ_n be defined by (39). It then follows from (11) by the d-optimality of T_n that $\liminf_n P_\delta(L_n < \lambda_n) \geq P(V_\delta < z) > \beta$. It follows hence from (40) that $a_n < \lambda_n$ for all sufficiently large n. The conclusion of the preceding paragraph and (39) now imply that $\limsup_n n^{-1/2} [\log \alpha(\phi_n) + n\nu] \leq z$; since z is arbitrary, it follows from (32) that (43) holds.

Remarks 9. A general form of Stein's lemma states that, in testing a singleton θ against δ, $\log \xi_n(\delta:\beta) = -nK(\delta,\theta) + o(n)$ provided only that (1), (2), and (3) hold with $K(\delta,\theta) < \infty$; see [BR], [BGZ], and Raghavachari (1983). The expansion (42) is evidently an extension (to composite Θ_0) and refinement (to second-order) of the general Stein lemma. Such extensions and refinements are more or less implicit in [BCL] and some of the references therein; we think the present explicit account might be of interest. It follows from Remarks 6 that (42) is valid for every β if Assumption 4 and (28) hold.

Remarks 10. In the present context, efficiency to first order, i.e. $\log \alpha(\phi_n) = -n\nu + o(n)$, does not necessarily imply efficiency to second order, even in very simple cases with V_δ a normal variable. To see this, suppose that n takes all values 1,2,... ; that $\mathbf{B}_n \subset \mathbf{B}_{n+1}$ for each n; that (42) holds and there exists a second-order efficient ϕ_n. For each n let m(n) be the positive integer such that $n - n^{1/2} < m(n) \leq n - n^{1/2} + 1$, and define $\phi_n^0 = \phi_{m(n)}$. Then $\log \alpha(\phi_n^0) = -n\nu + n^{1/2}(q + \nu) + o(n^{1/2})$, so ϕ_n^0 is first-order efficient but not second-order efficient. This demonstration does not, however, contradict statements in [BCL] and in Kallenberg (1983) that first order efficiency does imply second order efficiency since the statements cited concern critical functions which satisfy certain structural conditions. Cf. Remark 4 in Section 2.

Acknowledgment. The authors are grateful to Professor W. C. M. Kallenberg for many helpful comments and suggestions.

REFERENCES

Bahadur, R. R., Chandra, T. K., and Lambert, D. (1982). Some further properties of likelihood ratios on general sample spaces. Proc. Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, 1-19. Indian Statistical Institute, Calcutta.

Bahadur, R. R., Gupta, J. C., and Zabell, S. L. (1980). Large deviations, tests and estimates. Asymptotic Theory of Statistical Tests and Estimation (I. M. Chakravarti, Ed.), 33-64. Academic Press.

Bahadur, R. R., and Raghavachari, M. (1972). Some asymptotic properties of likelihood ratios on general sample spaces. Proc. Sixth Berkeley Symp. Math. Statist. Prob., I, 129-152. University of California Press.

Billingsley, P. (1968). Convergence of probability measures. John Wiley.

Fernique, X. (1974). Regularite des trajectoires des fonctions aleatoires Gaussiennes. Ecole d'Eté de Calcul des Probabilités de Saint-Flour-IV, Lecture Notes in Mathematics No. 480, 1-96. Springer-Verlag.

Kallenberg, W. C. M. (1981). Bahadur deficiency of likelihood ratio tests in exponential families. J. Mult. Anal., 11, No. 4, 506-531.

Kallenberg, W. C. M. (1983). Asymptotic efficiency and deficiency of tests. Proceedings of the 44th Session of the International Statistical Institute; Bulletin of the ISI, Vol. L, Bk. 2, 1173-1189.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley.

Raghavachari, M. (1983). On the computation of Hodges-Lehmann efficiency. A Festschrift for Erich L. Lehmann (Bickel, Doksum, and Hodges; Eds.), 367-378. Wadsworth.

Tsirel'son, V. S. (1975). The density of the distribution of the maximum of the Gaussian process. Theory Prob. Applications, 20, 847-856.