The Annals of Statistics
1980, Vol. 8, No. 6, 1382–1387

REMARKS ON BAHADUR OPTIMALITY OF CONDITIONAL TESTS

BY I. V. BASAWA

La Trobe University

It is customary to use a conditional test-statistic when testing for one parameter in a two-parameter exponential family model, the other parameter being considered as a nuisance parameter. In this note we discuss the Bahadur efficiency of such a condtional test and illustrate the results with examples from a branching process, a pure birth process, and a stable process.

1. Introduction. Let $X^n = (X_1, X_2, \cdots, X_n)$ be a vector of observations having a density $p_{X^n}(x^n; \theta, \eta)$ with respect to some product measure $\mu_n(x^n)$. We suppose that $(\theta, \eta) \in \Omega$, an open subset of R^2, R being the real line. The observations $(X_1, X_2, \cdots, X_n)$ are not necessarily identically distributed and they are allowed to be mutually dependent. Assume for the time being (see Theorem 2.3) that the density $p_{X^n}(x^n; \theta, \eta)$ belongs to an exponential family, viz.

$$p_{X^n}(x^n; \theta, \eta) = C_n(\theta, \eta) \exp[\theta U_n(x^n) + \eta V_n(x^n)] \tag{1.1}$$

where U_n and V_n are statistics and C_n does not depend on the observations. See Lehmann (1959) for a detailed description of the densities of the type (1.1). The two dimensional statistic (U_n, V_n) occurring in (1.1) is a minimal sufficient statistic for (θ, η).

Consider the problem of testing

$$H: \theta = \theta_0 \quad \text{against} \quad K: \theta > \theta_0$$

where η is a nuisance parameter in both the hypotheses. Now, the conditional density, $p_{U_n|V_n}(U_n; \theta)$, of U_n given $V_n = v_n$ is known to be free from the nuisance parameter η and in fact, it belongs to a one-parameter exponential family; thus

$$p_{U_n|V_n}(U_n; \theta) = C_n^*(v_n, \theta) \exp(\theta U_n) \tag{1.2}$$

where C_n^* depends on the observations only through v_n. Since (1.2) is free from η it is customary to conduct inference about θ using the density (1.2) rather than (1.1). We have for $\theta_1 > \theta_0$,

$$\frac{p_{U_n|V_n}(U_n; \theta_1)}{p_{U_n|V_n}(U_n; \theta_0)} = \frac{C_n^*(V_n, \theta_1)}{C_n^*(V_n, \theta_0)} \exp[(\theta_1 - \theta_0)U_n] \uparrow \text{ in } U_n.$$

It therefore follows by the Neyman-Pearson lemma that given V_n, a uniformly most powerful (UMP) size-α test of H against K is of the form

$$\begin{aligned} \phi_n(U_n, V_n) &= \begin{cases} 1 & \text{if} \quad U_n > k_n(V_n) \\ a_n(V_n) & \text{if} \quad u_n = k_n(V_n) \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1.3}$$

where $k_n(V_n)$ and $a_n(V_n)$ are determined by

$$E_{\theta_0}[\phi_n(U_n, V_n) \mid V_n = v_n] = \alpha.$$

The power function of this conditional test is

$$\beta_{\phi_n}(\theta \mid v_n) = E[\phi_n(U_n, V_n) \mid V_n = v_n] \tag{1.4}$$

Received July 1978; revised February 1979.

AMS 1970 *subject classification.* 62F05, 62M99, 62F20.

Key words and phrases. Exponential family, conditional tests, Bahadur efficiency, Galton-Watson process, pure birth process, stable process.

which is free from η but depends on v_n. Note, however, that the observation v_n is not available in advance and therefore one needs to view the conditional power function $\beta_{\phi_n}(\theta \mid V_n)$ as a random variable. One obvious way of dealing with this situation is to look at the average power function, $E_{\theta,\eta}[\beta_{\phi_n}(\theta \mid V_n)]$, as the criterion. This amounts to assessing the performance of the conditional test-statistic $U_n \mid V_n$ with reference to the unconditional density (1.1) rather than the conditional density (1.2). Such an unconditional argument leads to a justification of ϕ_n in (1.3) as a UMP unbiased size-α test (see Lehmann (1959), Chapter 4, Section 4, or Kendall and Stuart (1973), Chapter 23). However, it may be desirable to assess the conditional test-statistic $U_n \mid V_n$ directly with reference to the conditional density (1.2). Following Bahadur (1960, 1967) and Bahadur and Raghavachari (1971) we shall use the level (conditional) (see Definition 2.1 below) rather than the power as the underlying optimality criterion and discuss the Bahadur efficiency of the statistic $U_n \mid V_n$ in the following section. Although we have used a "discrete-time" formulation in Sections 1 and 2, an analogous "continuous-time" formulation is obvious. In Section 3 we briefly discuss some examples from discrete as well as continuous time models.

2. Bahadur optimality criterion. Suppose $X^n = (X_1, X_2, \cdots, X_n)$ has a density $p_{X^n}(x^n; \alpha, \beta)$ where the parameters α and β take values in open real intervals A and B respectively, and $(\alpha, \beta) \in \Omega = A \times B$. Suppose one is interested in testing $H: \alpha = \alpha_0$ treating β as a nuisance parameter. Let T_n and S_n be two statistics and suppose we are interested in assessing the performance of the conditional statistic $T_n \mid S_n$ for testing H for large n. Assume that large values of T_n correspond to the rejection of H. Let $h = (\alpha_0, \beta_0)$ and $k = (\alpha_1, \beta_1)$, $(\alpha_1 > \alpha_0)$ be typical elements of H and K respectively.

DEFINITION 2.1. The conditional *level attained* by the statistic $T_n \mid S_n$ is given by

$$Z_n(X^n) = \sup_{h \in \Omega_H} [P_{(h)}\{T_n \geq T_n(X^n) \mid S_n = S_n(X^n)\}] \tag{2.1}$$

where $P_{(h)}$ denotes the probability computed under the hypothesis $h = (\alpha_0, \beta_0)$, and $\Omega_H = \{(\alpha, \beta): \alpha = \alpha_0, \beta \in B\}$. Thus, $Z_n(X^n)$ is the largest probability under H of obtaining a value of T_n not less than the observed $T_n(X^n)$ conditional on $S_n = S_n(X^n)$. $Z_n(X^n)$ is also known as the (conditional) P-value of the statistic T_n given S_n.

Typically, $Z_n(X^n) \to 0$ almost surely (a.s.) under any fixed k. The rate at which this convergence obtains can be used as an optimality criterion.

Define

$$K_n(k, h) = a_n^{-1}(k) \log \frac{p_{X^n}(X^n; k)}{p_{X^n}(X^n; h)} \tag{2.2}$$

where $0 < a_n(k) \uparrow \infty$ is a suitably chosen sequence of real numbers (possibly depending on k) such that

$$K_n(k, h) \to K(k, h), \text{ a.s. under } k, \tag{2.3}$$

where $0 \leq K(k, h) < \infty$. *Let*

$$J(k) = \inf_{h \in \Omega_H} \{K(k, h)\}. \tag{2.4}$$

The following result is a generalization of that in Bahadur and Raghavachari (1971, Corollary 3):

THEOREM 2.1. *Using the notation introduced above we have*

$$\liminf_{n \to \infty} \{a_n^{-1}(k) \log Z_n(X^n)\} \geq -J(k), \text{ a.s. } under\ k \tag{2.5}$$

for any T_n and S_n.

REMARK. Bahadur and Raghavachari use the norm n in place of our $a_n(k)$. The more general norm $a_n(k)$ will be required in some of the examples we discuss in Section 3.

DEFINITION 2.2. (*Exact slope*). Suppose $-2a_n^{-1}(k) \log Z_n(X^n) \to s_k(T_n \mid S_n)$, a.s. under k, where $0 \le s_k(T_n \mid S_n) < \infty$. Then, $s_k(T_n \mid S_n)$ is called the exact slope of $T_n \mid S_n$ against k.

DEFINITION 2.3. (*Bahadur efficiency*). The statistic $T_n \mid S_n$ is said to be efficient for testing H against K if $s_k(T_n \mid S_n) = 2J(k)$.

REMARK. Note that in view of Theorem 2.1, $s_k(T_n \mid S_n) \le 2J(k)$ for any statistic $T_n \mid S_n$ and this inequality provides a natural motivation for Definition 2.3.

We now return to the case when X^n has a density of the form (1.1). In many examples satisfying (1.1) it is possible to effect a one-to-one transformation, (θ, η) to (θ, ξ_n) so that the reparametrized likelihood function, $L_n(X^n; \theta, \xi_n)$ can be factorized as

$$L_n(X^n; \theta, \xi_n) = L_n^c(U_n \mid V_n; \theta) L_n(V_n; \xi_n) \tag{2.6}$$

where the first factor corresponds to the conditional density of U_n given V_n and the second factor corresponds to the marginal density of V_n. Note that ξ is allowed to depend on n. (This does not affect our results.) The first factor on the right of (2.6) is free from ξ_n and the second factor does not depend on θ. We also assume that

$$(\theta, \xi_n) \in \Omega = \Theta \times \Phi_n. \tag{2.7}$$

The problem of testing $H: \theta = \theta_0$ against $K: \theta > \theta_0$ is unaffected by the above reparametrization. The following result then establishes the efficiency of the statistic $U_n \mid V_n$ in the sense of Definition 2.3.

THEOREM 2.2. *Suppose for the density of the type* (1.1) *a one-to-one reparametrization* $(\theta, \eta) \to (\theta, \xi_n)$, *exists such that* (2.6) *and* (2.7) *are satisfied. Then, the statistic* $U_n \mid V_n$ *is efficient in the sense of Definition* 2.3.

PROOF. Choose and fix an alternative parameter point $k = (\theta_1, \xi_{n1})$. Corresponding to this choice we may choose a null point $h_k = (\theta_0, \xi_{n1})$. Consider the problem of testing

$$H: p_{(\theta_0, \xi_{n1})} \quad \text{against} \quad K: p_{(\theta_1, \xi_{n1}), (\theta_1 > \theta_0)}.$$

With the notation of Bahadur and Raghavachari (1971, Section 3),

$$\begin{aligned} \Delta_n(X^n) &= a_n^{-1}(k) \log\{r_n(X^n)/\rho_n(X^n)\} \\ &= a_n^{-1}(k) \log\left\{\frac{L_n^c(U_n \mid V_n; \theta_1)}{L_n^c(U_n \mid V_n; \theta_0)}\right\} \\ &= K_n(k, h_k), \quad \text{since (2.6) obtains.} \\ &\to K(k, h_k) \quad \text{a.s. under } k. \end{aligned}$$

Now,

$$\begin{aligned} \hat{L}_n(X^n) &= p_{(\theta_0, \xi_{n1})}\{r_n \ge r_n(X^n) \mid V_n = V_n(X^n)\} \\ &= P_{(\theta_0, \xi_{n1})}\{U_n \ge U_n(X^n) \mid V_n = V_n(X^n)\} \end{aligned}$$

in virtue of (2.6) and (1.2).

A generalization of Theorem 3 of Bahadur and Raghavachari (1971) gives

$$a_n^{-1}(k) \log \hat{L}_n(X^n) \to -K(k, h_k) \quad \text{a.s. under } k.$$

Note, however, that since the conditional distribution of U_n given V_n is free from ξ_n, we have

$$\begin{aligned} Z_n(X^n) &= \sup_h [p_{(h)}\{U_n \ge U_n(X^n) \mid V_n = V_n(X^n)\}] \\ &= p_{(\theta_0, \xi_{n1})}\{U_n \ge U_n(X^n) \mid V_n = V_n(X^n)\} \\ &= \hat{L}_n(X^n). \end{aligned}$$

Therefore,

$$a_n^{-1}(k)\log Z_n(X^n) \to -K(k,h_k) \quad \text{a.s. under } k. \tag{2.8}$$

By Theorem 2.1, we must have

$$K(k, h_k) \le J(k).$$

But, by definition, $J(k) \le K(k, h_k)$, and hence we have equality $K(k, h_k) = J(k)$. This equality in conjunction with (2.8) shows that $U_n \mid V_n$ has the exact slope $2J(k)$ and hence it is efficient according to Definition 2.3. □

The result in Theorem 2.2 remains valid when the restriction to exponential family (1.1) is relaxed and also if (2.6) is weakened as follows:

THEOREM 2.3. *Suppose (U_n, V_n) is a minimal sufficient statistic for (θ, η) in the density $p_{X^n}(x^n; \theta, \eta)$ which admits a one-to-one reparametrization $(\theta, \eta) \to (\theta, \xi_n)$ such that the reparametrized likelihood function $L_n(X^n; \theta, \xi_n)$ can be factorized as*

$$L_n(X^n; \theta, \xi_n) = L_n^c(U_n \mid V_n; \theta) L_n(V_n; \theta, \xi_n) \tag{2.9}$$

where the first factor on the right is free from ξ_n and the second factor satisfies

$$a_n^{-1}(k)\log\frac{L_n(V_n; \theta_1, \xi_{n1})}{L_n(V_n; \theta_0, \xi_{n1})} \to 0 \quad \textit{under} \quad k \quad \text{a.s.} \tag{2.10}$$

Further, assume that

$$(\theta, \xi_n) \in \Omega = \Theta \times \Phi_n \tag{2.11}$$

and that $L_n(X^n; \theta, \xi_n)$ possesses a monotone likelihood ratio in U_n; more specifically,

$$\frac{L_n(X^n; \theta_1, \xi_{n1})}{L_n(X^n; \theta_0, \xi_{n1})} \uparrow \quad \textit{in} \quad U_n, (\theta_1 > \theta_0). \tag{2.12}$$

For testing $\theta = \theta_0$ against $\theta > \theta_0$ for the likelihood $L_n(X^n; \theta, \xi_n)$ satisfying (2.9) *to* (2.12), *the conditional statistic $U_n \mid V_n$ is efficient in the sense of Definition* 2.3.

This result can be proved as in Corollary 4 of Bahadur and Raghavachari and we omit the details. We shall now discuss some examples.

3. Examples.

EXAMPLE 1. (*Independent and identically distributed observations*). Let $X_1, \cdots X_n$ be independent $N(\mu, \sigma^2)$ variables where both μ and σ^2 are unknown. Consider the classical problem of testing $\sigma^2 = \sigma_0^2$ against $\sigma^2 > \sigma_0^2$. It is easily verified that

$$p_{X^n}(x^n; \theta, \eta) = C_n(\theta, \eta)\exp(\theta u_n + \eta v_n)$$

where $\theta = \sigma^{-2}$, $\eta = \mu\sigma^{-2}$, $u_n = -\frac{1}{2}\sum_1^n x_i^2$, and $v_n = \sum_1^n x_i$. We will show that the statistic $T_n \mid V_n$ is efficient according to Bahadur's criterion for testing $\theta = \theta_0$ against $\theta < \theta_0$ where $T_n = -2U_n$.

We may reparametrize (θ, η) to (θ, μ), the transformation being one-to-one. The likelihood function can now be written as

$$L_n(T_n, V_n; \theta, \mu) = L_n^c(T_n \mid V_n; \theta) L_n(V_n; \theta, \mu).$$

Note that while the first factor on the right of the above equation is free from the nuisance parameter μ the second factor, however, depends on θ. Choosing $a_n(k) = n$ it is easily verified that (2.10) is satisfied. Therefore, by Theorem 2.3 $T_n \mid V_n$ is efficient. The exact slope of $T_n \mid V_n$ is seen to be

$$\left\{\frac{(\theta_0 - \theta_1)}{\theta_1} - \log(\theta_0/\theta_1)\right\},$$

which, incidently, is free from the nuisance parameter.

EXAMPLE 2. (*Independent but not identically distributed observations*; *supercritical branching process*). Let X_j, $(j = 1, 2, \cdots, n)$, be independent Poisson random variables with $E(X_j) = \alpha\beta^j$ where $\alpha > 0$ and $\beta > 1$ are unknown parameters. Consider the problem of testing $\beta = \beta_0$ against $\beta > \beta_0$. The density of $X^n = (X_1, \cdots, X_n)$ then satisfies (1.1) with

$$\theta = \log\beta, \qquad \eta = \log\alpha, \qquad u_n = \textstyle\sum_1^n jx_j \quad \text{and} \quad v_n = \sum_1^n x_j.$$

Consider the one-one transformation $(\theta, \eta) \rightarrow (\theta, \xi_n)$ where

$$\xi_n = \left(\frac{e^{\theta+\eta}}{e^\theta - 1}\right)(e^{n\theta} - 1).$$

The reparametrized likelihood function of (θ, ξ_n) can be factorized as in (2.6). Note that V_n is a Poisson random variable with mean ξ_n. Choose $a_n(k) = n\xi_{n1}$ and verify that Theorem 2.2 is applicable. Thus, $U_n \mid V_n$ is efficient. Note, however, that the exact slope of $U_n \mid V_n$ is zero.

The above example is closely related to the supercritical Galton-Watson process with $Z_0 = 1$, $Z_1, Z_2, \cdots, Z_n$ being successive generation sizes and Z_1 distributed as a geometric random variable with mean β, $(\beta > 1)$. It is well known that for such a process $Z_n\beta^{-n}$ converges almost surely to a positive random variable W; in the present case W has a negative exponential distribution with mean unity. See Basawa and Scott (1976) for a discussion of a testing problem for this model. It is argued by Lauritzen (1976) that any asymptotic inference for β should be conducted conditionally on $W = w$, treating w as a nuisance parameter. It is known that conditionally on $W = w$, $X_j = (Z_j - Z_{j-1})$, $(j = 1, 2, \cdots)$, are independent Poisson random variables with means $(w(\beta - 1)/\beta)\beta^j$. Thus, taking $\alpha = w(\beta - 1)/\beta$ this problem reduces to Example 2 above.

EXAMPLE 3. (*Inhomogeneous Poisson process*; *pure birth process*). This is a continuous-time analogue of Example 2. Let $\{X(s), s \geq 0\}$ be a Poisson process with intensity $e^{\mu+\lambda t}$, $(\lambda > 0)$. Set $X(0) = 1$ and observe the process over the fixed interval $(0, t)$. It is seen that the likelihood function (see Keiding (1974)) belongs to the exponential family (1.1) with

$$L_t(\lambda, \mu) = C_t(\lambda, \mu)\exp[U(t)\lambda + V(t)\mu]$$

where $U(t) = tX(t) - S(t)$, $V(t) = X(t) - 1$, and $S(t) = \int_0^t X(s)\,ds$. Suppose we wish to test $\lambda = \lambda_0$ against $\lambda > \lambda_0$. We may reparametrize $(\lambda, \mu) \rightarrow (\lambda, \xi_t)$ where $\xi_t = e^\mu(e^{\lambda t} - 1)\lambda^{-1}$. We then obtain the factorization

$$L_t(\lambda, \xi_t) = L_t^c(U(t) \mid V(t); \lambda)L_t(V(t); \xi_t),$$

i.e., (2.6) is satisfied. We find that the continuous-time analogue of Theorem 2.2 is applicable and hence $U(t) \mid V(t)$ is optimal. Here we need to choose $a_t(k) = t\xi_{t1}$. It can be verified that the exact slope of $U(t) \mid V(t)$ is zero.

Example 3 is closely related to the pure birth process $\{Z(s), s \geq 0\}$ as follows. Let $\{Z(s), s \geq 0\}$ with $Z(0) = 1$ be a pure birth process with birth rate λ. It is well known that $Z(s)e^{-\lambda s}$ converges, as $s \rightarrow \infty$, almost surely to a negative exponential (unit mean) random variable W. As in the case of the supercritical Galton-Watson process in the previous example it can be argued that any asymptotic inference regarding λ should be conducted conditionally on $W = w$ (see Keiding (1974)) treating w as a nuisance parameter. It is known that conditionally on $W = w$, $\{Z(s), s \geq 0\}$ is an inhomogeneous Poisson process with $Z(0) = 1$ and intensity $w\lambda e^{\lambda t}$. Thus, setting $\mu = \log(w\lambda)$ this problem reduces to Example 3.

EXAMPLE 4. (*Additive stable process*). Let $\{X(s), s \geq 0\}$ be an additive nondecreasing stable process discussed by Basawa and Brockwell (1978). The Laplace transform of $X(s)$ is given by

$$E[\exp\{-\lambda X(s)\}] = \exp\left\{\frac{-s\alpha\lambda^\beta}{\Gamma(1-\beta)}\right\}, \qquad (\alpha > 0, 0 < \beta < 1),$$

where α and β are the unknown parameters. Suppose we wish to test $\beta = \beta_0$ against $\beta > \beta_0$ using a suitable realization of the process. Choose and fix $\epsilon > 0$. Suppose we observe $\{X(s)\}$ continuously over the fixed interval $(0, t)$ but record only jumps of size not less than ϵ. Let $\{(\tau_h, Y_h), h = 1, 2, \cdots, N(t)\}$ be the observed realization τ_h is the epoch at which hth jump $(\geq \epsilon)$ occurs and Y_h the jump-size $(Y_h \geq \epsilon)$; $N(t)$ is the total number of jumps in $(0, t)$ whose size is not less than ϵ. The likelihood function corresponding to this realization is seen to be of the type (1.1), i.e.,

$$L_t(\eta, \beta) = C_t(\beta, \eta) \exp[U(t)\beta + V(t)\eta]$$

where $\eta = \log(\alpha\beta\epsilon^{-\beta})$, $U(t) = -\sum_{j=1}^{N(t)} Z_j$, $Z_j = \log(Y_j/\epsilon)$, and $V(t) = N(t)$.

Let $T(t) = -U(t)$ and consider the conditional test-statistic $T(t) | V(t)$. We may reparametrize (β, η) to (β, ξ) where $\xi = e^{\eta}\beta^{-1}$. It is then seen that the factorization (2.6) obtains. Choosing $a_t(k) = t\xi_1$ we find that, by Theorem 2.2, $T(t) | V(t)$ is efficient.

It is not difficult to verify using equation (13) of Basawa and Brockwell that

$$\log\left\{\frac{L_t^c(U(t) | V(t); \beta_1)}{L_t^c(U(t) | V(t); \beta_0)}\right\} = V(t)\log(\beta_1/\beta_0) - (\beta_1 - \beta_0)T(t).$$

Now, under k, $E(V(t)) = t\xi_1$ (since $V(t)$ is Poisson with mean $t\xi$), and hence $a_t^{-1}(k)V(t)$ converges to unity almost surely under k. Note also that $a_t^{-1}(k)T(t)$ converges a.e. (k) to β_1^{-1} via the law of large numbers applied to the random sum $T(t)$; note that Z_j, $(j = 1, 2, \cdots,)$ are independent identically distributed exponential random variables with mean β^{-1}. The exact slope of the statistic $T(t) | V(t)$ is then seen to be

$$2\{\log(\beta_0/\beta_1) + (\beta_1 - \beta_0)/\beta_1\}.$$

This expression can be obtained alternatively by noting that conditional on $V(t)$, $T(t)$ is a gamma random variable mean $V(t)/\beta$ and index $V(t)$.

Acknowledgment. I am greatly indebted to a referee for valuable and constructive comments which led to some improvements both in content and presentation (especially in the proof of Theorem 2.2).

REFERENCES

[1] BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295.
[2] BAHADUR, R. R. (1967). Rates of convergence of estimates and test-statistics. *Ann. Math. Statist.* **38** 303–324.
[3] BAHADUR, R. R. and RAGHAVACHARI, M. (1971). Some asymptotic properties of likelihood ratios on general sample spaces. In *Proc. 6th Berkeley Symp. Math. Statist. Probability* **1** 129–152. Univ. California Press.
[4] BASAWA, I. V. and BROCKWELL, P. J. (1978). Inference for gamma and stable processes. *Biometrika* **65** 129–33.
[5] BASAWA, I. V. and SCOTT, D. J. (1976). Efficient tests for branching processes. *Biometrika* **63** 531–536.
[6] JOHANSEN, S. (1976). Two notes on conditioning and ancillarity. Preprint, Univ. Copenhagen.
[7] KEIDING, N. (1974). Estimation in the birth process. *Biometrika* **61** 71–80.
[8] KENDALL, M. G. and STUART, A. (1973). *The Advanced Theory of Statistics, Vol. 2, 3rd ed.* Griffin, London.
[9] LAURITZEN, S. L. (1976). Projective statistical fields. Preprint, Univ. Copenhagen.
[10] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

DEPARTMENT OF MATHEMATICAL STATISTICS
LA TROBE UNIVERSITY
BUNDOORA, VICTORIA 3083
AUSTRALIA