The Annals of Statistics
1984, Vol. 12, No. 4, 1522–1527

BAHADUR OPTIMALITY OF SEQUENTIAL EXPERIMENTS FOR EXPONENTIAL FAMILIES[1]

BY STAVROS KOUROUKLIS

The Pennsylvania State University

A theorem of Bahadur on the asymptotic optimality of the likelihood ratio statistic has been extended to sequential analysis by Berk and Brown (1978) in the context of testing one-sided hypotheses about the mean of a normal distribution with known variance. In this work, Bahadur's theorem is extended to sequential analysis for general hypotheses about the parameters of an exponential family of distributions. Specifically, it is shown that, under certain conditions, modifications of the likelihood ratio statistic analogous to those exhibited by Berk and Brown (1978) in the above normal context are optimal for any family of stopping times approaching ∞. These results indicate that Bahadur efficiency has a limited impact in sequential analysis.

1. Introduction. The notion of Bahadur efficiency and the optimality (under certain conditions) of the likelihood ratio statistic (l.r.s.) in the sense of Bahadur in a nonsequential context have long been known. Lately, Berk and Brown (1978) extended this notion to sequential analysis, and showed that, in contrast to the nonsequential case, the l.r.s. for testing onesided hypotheses about the mean of a normal distribution with known variance is not always optimal, i.e., there is a family of stopping times for which the l.r.s. is nonoptimal. However, they obtained in the above normal context modifications of the l.r.s. which are optimal for any family of stopping times approaching ∞. In the same work the authors exhibited a class of stopping times for which the sample mean is optimal, but showed by an example that the sample mean is not optimal for an arbitrary family of stopping times.

In this paper we generalize work of Berk and Brown (1978) to exponential families of distributions. For these models we show that, under certain regularity conditions, modifications of the l.r.s. analogous to those considered by Berk and Brown (1978) in the normal case are optimal for any family of stopping times approaching ∞. See Theorem 3.2.

The results of Berk and Brown (1978) as well as ours prove that there are infinitely many sequential tests (yielded either by the same statistic and different stopping times or different statistics and the same stopping time or different statistics and different stopping times) which perform equally well (in fact are optimal) in the sense of Bahadur, and hence one cannot conclude which of them is "better". Thus, the notion of Bahadur efficiency has a limited impact in sequential testing. The phenomenon, however, of many tests being Bahadur optimal has also been encountered in nonsequential analysis (see e.g. Chandra and Ghosh, 1978); there, the introduction of the notion of Bahadur deficiency by

Received September 1983.
[1] Research supported in part by NSF Grant to Prof. R. H. Berk.
AMS 1980 *subject classifications.* Primary 62L10, 62F05; secondary 60F10
Key words and phrases. Sequential test, Bahadur efficiency, exponential family, large deviations.

Kallenberg (1981) has provided further information about the performance of such tests. Perhaps, what is next to be sought is some sort of extension of this notion to sequential analysis.

In one dimension our result (Theorem 3.2) is obtained by using a lemma of Kallenberg (1978) (see Lemma 1.1 in Kourouklis, 1984), while in higher dimensions by using a large deviation result of Kourouklis (1984). Theorem 3.2 can handle any univariate or multivariate normal testing problem. In one dimension it applies to any testing problem.

Section 2 contains a synopsis of Berk's and Brown's theory on the extension of Bahadur efficiency to sequential analysis. Section 3 contains our assumptions and result.

2. Sequential Bahadur efficiency. Let $X_1, X_2, \cdots$ be a data sequence of i.i.d. abstract random variables with family of distributions $\mathscr{P} = \{P_\omega: \omega \in \Omega\}$. To simplify notation, throughout, P_ω will denote both the joint distribution of the data sequence and the marginal distribution of $(X_1, \cdots, X_n)$ for any $n \geq 1$. When unclear, it will be explicitly stated which of these two meanings P_ω stands for. We consider sequential tests of H_0: $\omega \in \Omega_0$ vs. H_1: $\omega \in \Omega_1$, where Ω_0, Ω_1 are disjoint nonempty proper subsets of Ω. Usually, $\Omega_1 = \Omega \sim \Omega_0$, although we do not require such an assumption here.

Let $\{T_n = T_n(X_1, \cdots, X_n): n \geq 1\}$ be a sequence of real-valued test statistics for testing H_0 vs. H_1, where T_n is measurable with respect to $\mathscr{B}(X_1, \cdots, X_n)$, the σ-field generated by $X_1, \cdots, X_n$. Let also $\{N_a\}$ be a family of stopping times indexed by a real parameter a. We assume $P_\omega(N_a < \infty) = 1$, and for asymptotic considerations we require that for all $\omega \in \Omega$, the limit in probability, $P_\omega - \lim_a N_a = \infty$. Throughout, limits on a are taken as $a \to \infty$. This requirement on the stopping times simply reflects the nonsequential requirement that (for asymptotic theory) the sample size should approach ∞. The stopped value of the sequence $\{T_n\}$ for the stopping time N_a is denoted by $T(a)$. Assuming that large values are significant, the attained level of $T(a)$ is defined to be

$$L(a) = H(T(a)), \quad \text{where} \quad H(x) = \sup\{P_\omega(T(a) \geq x): \omega \in \Omega_0\}.$$

The Kullback-Leibler information number for $\omega, \partial \in \Omega$ is defined by

$$K(\omega, \partial) = \int \log\left(\frac{dP_\omega}{dP_\partial}\right) dP_\omega \tag{2.1}$$

if $P_\omega \ll P_\partial$; otherwise $K(\omega, \partial) = \infty$. Here, P_ω, P_∂ are meant to be distributions of X_1. We also let

$$K(\omega, \Omega_0) = \inf\{K(\omega, \partial): \partial \in \Omega_0\}. \tag{2.2}$$

Extending a theorem of Raghavachari (1970) to sequential analysis, Berk and Brown (1978) showed that $K(\omega, \Omega_0)$ provides an upper bound for limits in probability of $-\log L(a)/N_a$. Accordingly, the family $\{T(a)\}$ is called optimal at $\omega \in \Omega$ whenever this upper bound is attained.

DEFINITION. Let $L(a)$ be the attained level of $T(a)$, the stopped value of $\{T_n\}$

for the stopping time N_a. The family $\{T(a)\}$ is called optimal at $\omega \in \Omega$ if

$$P_\omega - \lim_a[-\log L(a)/N_a] = K(\omega, \Omega_0).$$

With this definition of optimality and in the context of testing $\omega \le 0$ vs. $\omega > 0$ when the data sequence is normal $N(\omega, 1)$, Berk and Brown (1978) showed that modifications of the l.r.s. (which, here, is strictly increasing function of $S_n/n^{1/2}$) of the form $S_n/n^{1/2} - c_n$, where $S_n = X_1 + \cdots + X_n$ and $\{c_n\}$ is a sequence of constants satisfying certain conditions, stopped by any family of stopping times are optimal at all $\omega > 0$.

3. Assumptions and result. Let $\mathscr{P} = \{P_\omega: \omega \in \Omega\}$ denote a k-dimensional natural exponential family of distributions with densities (at x)

$$dP_\omega/d\nu = \exp\{\omega' x - c(\omega)\}, \quad x \in R^k, \quad \omega \in \Omega,$$

with respect to a σ-finite measure ν on $\mathscr{B}(R^k)$. Here $'$ denotes transpose. Ω is the natural parameter space, i.e.,

$$\Omega = \left\{\omega \in R^k: \exp\{c(\omega)\} = \int \exp\{\omega' x\}\, d\nu(x) < \infty\right\}.$$

Throughout this paper we assume that Ω is an open subset of R^k. For $\partial \in \Omega$ consider the log likelihood ratio function

$$\varphi_\partial(x) = \sup\{(\omega - \partial)'x - c(\omega) + c(\partial): \omega \in \Omega\}, \quad x \in R^k.$$

Let now $X_1, X_2, \cdots$ be a sequence of i.i.d. random vectors in R^k with family of distributions $\mathscr{P}$. Set $\bar{X}_n = \sum_{i=1}^n X_i/n$, $n \ge 1$. The likelihood ratio statistic for testing H_0: $\omega \in \Omega_0$ vs. H_1: $\omega \in \Omega_1$ is defined to be

$$\ell_n = \frac{\sup\{\exp[n\omega'\bar{X}_n - nc(\omega)]: \omega \in \Omega\}}{\sup\{\exp[n\omega'\bar{X}_n - nc(\omega)]: \omega \in \Omega_0\}}, \quad n \ge 1.$$

We take, by convention, $\infty/\infty = 1$. Note that for $\Omega_0 = \{\partial\}$, $\ell_n = \exp\{n\varphi_\partial(\bar{X}_n)\}$. The following well-known lemma plays an important role in the sequel; its proof is omitted.

LEMMA 3.1. *For each $\omega \in \Omega$, $\log \ell_n/n \to K(\omega, \Omega_0)$ w.p.1 $[P_\omega]$ as $n \to \infty$. We now require that for $k \ge 2$ Assumptions 3.1–3.3 of Kourouklis (1984) hold with $w \in \Omega_0$, and let m, β be as in Assumptions 3.2 and 3.3 respectively.*

We set $k_0 = \beta k(k-1)$. We also make the following assumption (again for $k \ge 2$).

ASSUMPTION 3.4. Either the distribution of ℓ_n or the distribution of $\varphi_\partial(\bar{X}_n)$ under P_∂ do not depend on ∂, for all $\partial \in \Omega_0$ and all $n \ge m$.

For $k = 1$ we require no assumption whatsoever and set $m = 1$, $k_0 = 0$. Next, for $k \ge 1$ we let $n_1 = \max(m - 1, k_0)$. For $k \ge 2$, $\partial \in \Omega_0$, $\varepsilon > 0$, $0 < \tau < 1$, $n > n_1$, by Assumption 3.4 and Theorem 3.2 of Kourouklis (1984) we have

$$P_\partial((n-k_0)\log \ell_n/n > \varepsilon) \le P_w((n-k_0)\varphi_w(\bar{X}_n) > \varepsilon) \le c(\tau, w)n^{k(k-1)}\exp\{-\tau\varepsilon\},$$

where $c(\tau, w)$ is a constant.

For $k = 1$, $\partial \in \Omega_0$, $\varepsilon > 0$, $0 < \tau < 1$, $n > n_1$, by Lemma 1.1 in Kourouklis (1984) we have

$$P_\partial((n-k_0)\log \ell_n/n > \varepsilon) \le P_\partial((n-k_0)\varphi_\partial(\bar{X}_n) > \varepsilon) \le 2\exp\{-\tau\varepsilon\}.$$

Hence, setting $c(\tau) = 2$ or $c(\tau, w)$ according as $k = 1$ or $k \ge 2$ we have

$$P_\partial((n-k_0)\log \ell_n/n > \varepsilon) \le c(\tau)n^{k(k-1)}\exp\{-\tau\varepsilon\}, \tag{3.1}$$

for all $\partial \in \Omega_0$, $\varepsilon > 0$, $0 < \tau < 1$, $n > n_1$, $k \ge 1$.

We are now in position to state our result.

THEOREM 3.2. *Suppose that the above assumptions hold. Let*

$$T_n = a_n \log \ell_n - c_n, \quad \textit{if} \quad n > n_1$$
$$= 0, \qquad \textit{otherwise},$$

where $0 < a_n < \alpha(n-k_0)/n$ *for* $n > n_1$, $a_n \to \alpha$, *and for some* $\delta > 0$ *and* $0 < \tau_0 < 1$, $\alpha(1 + k(k-1) + \delta)\log n/\tau_0 \le c_n = o(n)$. *Then for any family* $\{N_a\}$ *of stopping times for which* $P_\omega - \lim_a N_a = \infty$, *the family* $\{T(a)\}$ *of the stopped values of the sequence* $\{T_n\}$ *is optimal at* ω, *for all* $\omega \in \Omega \sim \Omega_0$.

PROOF. Let $\partial \in \Omega_0$, $x > 0$, $\tau_0 \le \tau < 1$. Using (3.1) we obtain

$$\begin{aligned}
&P_\partial(T(a) \ge x)\\
&\quad \le \textstyle\sum_{n=1}^\infty P_\partial(T_n \ge x) = \sum_{n=n_1+1}^\infty P_\partial(T_n \ge x) = \sum_{n=n_1+1}^\infty P_\partial(a_n\log \ell_n - c_n \ge x)\\
&\quad \le \textstyle\sum_{n=n_1+1}^\infty P_\partial((n-k_0)\log \ell_n/n \ge c_n/\alpha + x/\alpha)\\
&\quad \le \exp\{-\tau x/\alpha\}c(\tau) \textstyle\sum_{n=n_1+1}^\infty n^{k(k-1)}\exp\{-\tau c_n/\alpha\}\\
&\quad \le \exp\{-\tau x/\alpha\}c(\tau) \textstyle\sum_{n=n_1+1}^\infty 1/n^{1+\delta} = c_1(\tau, \delta)\exp\{-\tau x/\alpha\}.
\end{aligned}$$

Hence,

$$\sup\{P_\partial(T(a) \ge x) : \partial \in \Omega_0\} \le c_1(\tau, \delta)\exp\{-\tau x/\alpha\}, \quad \text{all} \quad x > 0. \tag{3.2}$$

Let now $\omega \in \Omega \sim \Omega_0$. We may assume that $K(\omega, \Omega_0) > 0$, since otherwise the theorem is trivially true by Theorem 2.1 of Berk and Brown (1978). Since $P_\omega - \lim_a N_a = \infty$, Lemma 3.1 entails $P_\omega - \lim_a[T(a)/N_a] = \alpha K(\omega, \Omega_0)$. Letting $L(a)$ denote the attained level of $T(a)$, it follows from (3.2) that

$$-\log L(a) \ge \tau T(a)/\alpha - \log c_1(\tau, \delta) \quad \text{on} \quad (T(a) > 0),$$

and hence

$$P_\omega - \liminf_a[-\log L(a)/N_a] \ge \tau K(\omega, \Omega_0) \quad \text{w.p.1 } [P_\omega] \quad \text{since} \quad P_\omega(T(a) > 0) \to 1.$$

Letting $\tau \to 1$ we obtain $P_\omega - \liminf_a[-\log L(a)/N_a] \ge K(\omega, \Omega_0)$ w.p.1 $[P_\omega]$. In view of Theorem 2.1 of Berk and Brown (1978), we conclude

$$P_\omega - \lim_a[-\log L(a)/N_a] = K(\omega, \Omega_0),$$

i.e., $\{T(a)\}$ is optimal at ω. □

REMARK 1. For one-dimensional exponential families, Theorem 3.2 covers any testing problem about the natural parameter ω.

REMARK 2. Suppose that Assumptions 3.1–3.3 of Kourouklis (1984) are satisfied (with w not necessarily in Ω_0 as we require above) and Assumption 3.4 holds with the distribution of $\varphi_\partial(\bar{X}_n)$ under P_∂ independent of ∂ for all $\partial \in \Omega$ (rather than all $\partial \in \Omega_0$). Then, it follows from the method of the above proof that Theorem 3.2 can handle any testing problem about the natural parameter ω. This situation occurs in the case of normal data. That Assumptions 3.1–3.3 hold is shown in Kourouklis (1984), while that $\varphi_\partial(\bar{X}_n)$ has distribution under P_∂ independent of ∂ for all $\partial \in \Omega$ follows from invariance of the problem. Hence, Theorem 3.2 applies to any univariate or multivariate normal testing problem.

REMARK 3. Two nonnormal examples for which the regularity conditions of Theorem 3.2 are satisfied are testing that the shape parameter of a gamma distribution has a specified value and testing equality of the parameters of two independent negative exponential distributions. Proofs are given in Kourouklis (1981).

REMARK 4. In the case of multinomial distribution, which does not satisfy Assumption 3.1 of Kourouklis (1984), one can still obtain a result analogous to that of Theorem 3.2 (for any testing problem) by using the bound for $P_\eta(\varphi_\eta(\bar{X}_n) > \varepsilon)$ given in Remark 1 of Kourouklis (1984) and proceeding as in the proof of Theorem 3.2. Here $n_1 = 0$, $k_0 = 0$, $0 < a_n \le \alpha$, $a_n \to \alpha$ and for some $\delta > 0$ $\alpha(k + \delta)\log n \le c_n = o(n)$.

REMARK 5. When the data vectors consist of independent components, one need not have to verify Assumptions 3.1–3.4 but simply use the bound for $P_\partial(\varphi_\partial(\bar{X}_n) > \varepsilon)$ given in Remark 2 of Kourouklis (1984) and proceed as in the proof of Theorem 3.2. Here $n_1 = 0$, $k_0 = 0$, $0 < a_n \le \alpha$, $a_n \to \alpha$, and for some $\delta >$ 0 and $0 < \tau_0 < 1$, $\alpha(1 + \delta)\log n/\tau_0 \le c_n = o(n)$. Note that in this case there is no restriction on the testing problems that can be handled.

REMARK 6. When Ω is finite one may choose $0 \le c_n = o(n)$, allow a_n and c_n to depend on the data, and need not have $c_n \to \infty$. In this case too, $n_1 = 0$, $k_0 = 0$.

Acknowledgements. The author would like to express his gratitude to Prof. R. H. Berk, his thesis advisor, for many helpful discussions throughout the preparation of his Ph.D. dissertation of which this is a part. The author also expresses his sincere thanks to the editor, a referee, and an associate editor for their suggestions.

REFERENCES

BERK, R. H. and BROWN, L. D. (1978). Sequential Bahadur efficiency. *Ann. Statist.* **6** 567–581.

CHANDRA, T. K. and GHOSH, J. K. (1978). Comparison of tests with same Bahadur efficiency *Sankhyā Ser. A* **40** 253–277.

KALLENBERG, W. C. M. (1978). Asymptotic optimality of likelihood ratio tests in exponential families. *Math. Centre Tracts 77.* Amsterdam.

KALLENBERG, W. C. M. (1981). Bahadur deficiency of likelihood ratio tests in exponential families. *J. Multivariate Anal.* **11** 506–531.

KOUROUKLIS, S. (1981). Bahadur optimality of sequential experiments for exponential families. Doctoral dissertation, Rutgers University, New Brunswick.

KOUROUKLIS, S. (1984). A large deviation result for the likelihood ratio statistic in exponential families. *Ann. Statist.* **12** 1510–1521.

RAGHAVACHARI, M. (1970). On a theorem of Bahadur on the rate of convergence of test statistics. *Ann. Math. Statist.* **41** 1695–1699.

THE PENNSYLVANIA STATE UNIVERSITY
DEPT. OF STATISTICS
219 POND LABORATORY
UNIVERSITY PARK, PENNSYLVANIA 16802