*The Annals of Statistics*
1984, Vol. 12, No. 4, 1510–1521

# A LARGE DEVIATION RESULT FOR THE LIKELIHOOD RATIO STATISTIC IN EXPONENTIAL FAMILIES[1]

BY STAVROS KOUROUKLIS

*The Pennsylvania State University*

In this paper we consider exponential families of distributions and obtain under certain conditions a uniform large deviation result about the tail probability $P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon)$, $\varepsilon > 0$, where $\partial$ is the natural parameter and $\phi_\partial(\bar{X}_n)$ is the log likelihood ratio statistic for testing the null hypothesis $\{\partial\}$. The technique involves approximating certain convex compact sets in $R^k$ by polytopes, then estimating the probability contents of associated closed half-spaces, and counting the number of these half-spaces. Some examples are given, among them the multivariate normal distribution with unknown mean vector and covariance matrix.

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

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

with respect to a $\sigma$-finite measure $\nu$ on $\mathscr{B}(R^k)$. We say that $\nu$ generates the family $\mathscr{P}$. Here, $'$ denotes transpose. $\Omega$ is the natural parameter space, i.e.,

$$\Omega = \left\{\omega \in R^k : \exp\{c(\omega)\} = \int \exp\{\omega' x\}\, d\nu(x) < \infty\right\}. \tag{1.1}$$

Throughout this paper we assume that $\Omega$ is an open subset of $R^k$. Without loss of generality we may also suppose that $\nu$ is not supported on a flat.

For $\partial \in \Omega$ consider the log likelihood ratio function

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

Let now $X_1, X_2, \cdots$ be a sequence of i.i.d. random vectors in $R^k$ with family of distributions $\mathscr{P}$, and $\bar{X}_n = \sum_{i=1}^n X_i/n$. Then, $\phi_\partial(\bar{X}_n) = \log \ell_n/n$, $n \geq 1$, where $\ell_n$ is the likelihood ratio statistic for testing $H_0: \omega = \partial$ vs. $H_1: \omega \neq \partial$.

In this paper we study the tail probability $P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon)$, $\varepsilon > 0$. For one-dimensional exponential families, i.e., for $k = 1$, Kallenberg (1978, page 16) has proved the following lemma.

LEMMA 1.1. *Let $\varepsilon > 0$. Then, given $\partial \in \Omega$,*

$$P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon) \leq 2e^{-n\varepsilon}, \quad \textit{for all} \quad n \geq 1.$$

Received September 1983; revised May 1984.

[1] Research supported in part by NSF Grant to Prof. R. H. Berk.

*AMS* 1980 *subject classifications.* Primary 60F10, 62F03; secondary 52A20, 52A25.

*Key words and phrases.* Large deviations, exponential families, convexity, polytope.

The main result presented here may be considered as a multidimensional analog of Lemma 1.1. Specifically, we prove below that, for $k \geq 2$, under certain conditions

$$P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon) \leq c(\tau, \partial) n^{k(k-1)} \exp\{-\tau(n - k_0)\varepsilon\} \tag{1.2}$$

for all $\varepsilon > 0$, $0 < \tau < 1$, $n > n_1$, where the constants $n_1$, $k_0$, $c$ do not depend on $\varepsilon$. Relevant asymptotic results appearing in the literature include Efron and Truax (1968), Woodroofe (1978), Kallenberg (1978). Our result is cruder than those, but unlike those, holds uniformly in $\varepsilon \in (0, \infty)$. Besides, a version of (1.2) (see (3.13)) allows us to generalize uniform large deviation results in Hoeffding (1965) and in Herr (1967) to multidimensional exponential families (see Theorem (3.3)).

An interesting application of (1.2) is given in Kourouklis (1984). There the author uses (1.2) to obtain Bahadur optimal statistics in sequential context, thus generalizing work of Berk and Brown (1978). It can also be applied to nonsequential analysis to establish Bahadur optimality of the likelihood ratio statistic in both the one-sample and the multi-sample problems with data from exponential family (families) (see Kourouklis, 1981).

Section 2 contains some properties of exponential families we will need in the sequel. Section 3 contains our assumptions and the main result. In Section 4 we present some examples for which these assumptions are satisfied.

**2. Preliminaries.** In this section we collect some well known properties pertaining to exponential families. Details are given in Berk (1972), Barndorff-Nielsen (1978), and Rockafellar (1970).

The function $c$ defined in (1.1) is closed, strictly convex and infinitely differentiable on $\Omega$ with gradient $\dot{c}(\omega) = E_\omega X$ (expectation of $X$) and positive definite hessian $\ddot{c}(\omega) = \mathrm{Cov}_\omega X$ (covariance of $X$), where $X \sim P_\omega$, $\omega \in \Omega$.

Consider now the conjugate function of $c$, i.e.,

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

and let

$$B = \{x \in R^k: \phi(x) < \infty\}.$$

$\phi$ plays an important role in the sequel as well as in the theory of maximum likelihood estimation (see Berk, 1972). General theory on conjugate functions is contained in Rockafellar (1970). $\phi$ is closed, convex, and its conjugate is $c$. It is also differentiable on int$B$ (interior of $B$) and the restriction of $\dot{\phi}$ on $\dot{c}(\Omega)$ is the inverse of $\dot{c}$. Moreover, if $B$ is open (an assumption we require in the next section), from Theorems 9.1 and 9.2 of Barndorff-Nielsen (1978) it follows $B = \dot{c}(\Omega)$, so that $\dot{c}: \Omega \to B$ is 1-1 and onto with inverse $\dot{\phi}: B \to \Omega$.

Similarly, $\phi_\partial$ is the conjugate of $c(\omega + \partial) - c(\partial)$, hence it is a closed convex function. The set of finiteness of $\phi_\partial$ is also $B$, and $\phi_\partial$ has a minimum 0 attained at $x = \dot{c}(\partial)$.

We close this section by defining the Kullback–Leibler information number for $\omega, \partial \in \Omega$ by

$$K(\omega, \partial) = \int \log\left(\frac{dP_\omega}{dP_\partial}\right) dP_\omega = (\omega - \partial)'\dot{c}(\omega) - c(\omega) + c(\partial). \tag{2.1}$$

**3. Assumptions and main result.** In this section we consider multidimensional natural exponential families of distributions.

ASSUMPTION 3.1. $B$ is open.

ASSUMPTION 3.2. There is a positive integer $m$, such that $\bar{X}_m \in B$ with probability one with respect to some (and hence every) member of $\mathscr{P}$. Note that this implies $\bar{X}_n \in B$ with probability one for all $n \geq m$. It is implicitly meant that $m$ is the minimum possible.

For $\partial \in \Omega$ and $\varepsilon > 0$, we let

$$I_{\varepsilon,\partial} = \{\omega \in \Omega: K(\omega, \partial) \leq \varepsilon\}. \tag{3.1}$$

Below and throughout, det, tr denote determinant, trace respectively.

ASSUMPTION 3.3. For some $w \in \Omega$, there are constants $\beta = \beta(w) \geq 0$, $\gamma = \gamma(w) > 0$ and a real-valued Borel-measurable function $f(t) = \exp\{o(t)\}$ as $t \to \infty$ depending in general on $w$ and bounded on any finite interval $(0, \delta)$, such that for all $\varepsilon > 0$

$$\inf\{\det \ddot{c}(\omega): \omega \in I_{\varepsilon,w}\} \geq \gamma \exp\{-\beta\varepsilon\}$$

and

$$\sup\{\operatorname{tr} \ddot{c}(\omega): \omega \in I_{\varepsilon,w}\} \leq f(\varepsilon).$$

Assumptions 3.1 and 3.2 are equivalent to the existence of the maximum likelihood estimator of $\omega \in \Omega$ with probability one for $n \geq m$, as can be seen by Corollary 9.6 of Barndorff-Nielsen (1978). Assumption 3.1 is further discussed in Remark 1 following the proof of Theorem 3.2. A referee has noted that if the distribution of $\bar{X}_m$ is dominated by Lebesgue measure then both Assumptions 3.1 and 3.2 are satisfied, and if $B$ in Assumption 3.2 is replaced by int$B$ then Assumptions 3.1 and 3.2 are equivalent to each other. Assumption 3.3 facilitates the calculation of certain integrals. Typically, $f$ is a polynomial. In making this assumption, we were motivated by the two-parameter normal model $N(\mu, \sigma^2)$ where the corresponding bounds for the infimum and the supremum are asymptotically sharp (see Kourouklis, 1981, Example 1).

For $\partial \in \Omega$ and $\varepsilon > 0$ consider the level set

$$A_{\varepsilon,\partial} = \{x \in R^k: \phi_\partial(x) \leq \varepsilon\} \subset B. \tag{3.2}$$

Since $\phi_\partial$ is convex, so is $A_{\varepsilon,\partial}$. By Theorem 5.20 of Barndorff-Nielsen (1978), $A_{\varepsilon,\partial}$ is a compact neighborhood of $\dot{c}(\partial)$ (recall that $\phi_\partial$ has minimum 0 attained at $x = \dot{c}(\partial)$). Under Assumption 3.1, $\phi_\partial$ is continous on $B$ and hence $\phi_\partial(x) = \varepsilon$ for

$x \in bdA_{\varepsilon,\vartheta}$ (the boundary of $A_{\varepsilon,\vartheta}$). Moreover, it can be easily shown that $I_{\varepsilon,\vartheta} = \dot{\phi}(A_{\varepsilon,\vartheta})$. Hence, $I_{\varepsilon,\vartheta}$ is also compact.

The following lemma is well known and therefore its proof is omitted. We only note here that it also holds for $k = 1$ and in that case a minor generalization of the lemma yields an alternative proof of Lemma 1.1.

LEMMA 3.1. *Suppose that Assumption* 3.1 *holds. Consider* $\varepsilon > 0$, $\vartheta \in \Omega$ *and let* $G$ *be the closed half-space determined by a supporting hyperplane of* $A_{\varepsilon,\vartheta}$ *and not containing* $A_{\varepsilon,\vartheta}$. *Then,* $P_\vartheta(\bar{X}_n \in G) \le \exp\{-n\varepsilon\}$, $n \ge 1$.

Theorem 3.2 below is our main result. The type of argument used is related to that of Borovkov (1968) and Efron and Truax (1968).

THEOREM 3.2. *Suppose that Assumptions* 3.1–3.3 *hold and let*

$$k_0 = \beta k(k-1), \quad n_1 = \max(m-1, k_0).$$

*Then, given* $\tau$, $0 < \tau < 1$, $E_w \exp\{\tau(n-k_0)\phi_w(\bar{X}_n)\}$ *is finite for all* $n > n_1$ *and*

$$E_w \exp\{\tau(n-k_0)\phi_w(\bar{X}_n)\} = O(n^{k(k-1)}). \tag{3.3}$$

*Moreover, for some constant* $c(\tau, w)$,

$$P_w(\phi_w(\bar{X}_n) > \varepsilon) \le c(\tau, w) n^{k(k-1)} \exp\{-\tau(n-k_0)\varepsilon\}, \tag{3.4}$$

*for all* $\varepsilon > 0$, $0 < \tau < 1$, $n > n_1$.

PROOF. Consider the level sets $A_{\delta,w}$, $A_{\varepsilon,w}$, where $0 < \delta < \varepsilon$. Note that $A_{\delta,w} \subset A_{\varepsilon,w}$ and $d = \inf\{\|x - y\|: x \in bdA_{\delta,w}, y \in bdA_{\varepsilon,w}\} > 0$, since on the boundaries the value of $\phi_w$ is $\delta$, $\varepsilon$ respectively. By Proposition 5.6 of the Appendix, there is a polytope $\Pi$, $A_{\delta,w} \subset \Pi \subset A_{\varepsilon,w}$, which is the intersection of at most $c_0[(r+1)/\delta_1]^{k(k-1)}$ closed half-spaces, where $r$ is the circumradius of $A_{\varepsilon,w}$, $\delta_1 = \min(\pi/8, d/2)$, and $c_0$ a (positive) constant depending only on $k$. Let $F_i$, $i = 1, \cdots, p$ be the open half-spaces such that $\Pi = \cap_{i=1}^p F_i^c$. Since $A_{\delta,w} \subset \Pi$, each $F_i$ is contained in a closed half-space, $G_i$ say, determined by a hyperplane supporting $A_{\delta,w}$ and parallel to the hyperplane determining $F_i$. By Lemma 3.1, $P_w(\bar{X}_n \in F_i) \le P_w(\bar{X}_n \in G_i) \le \exp\{-n\delta\}$, for all $i = 1, \cdots, p$ and $n \ge 1$. Since $\Pi \subset A_{\varepsilon,w}$,

$$\begin{aligned} P_w(\phi_w(\bar{X}_n) > \varepsilon) &= P_w(\bar{X}_n \in A_{\varepsilon,w}^c) \le pP_w(\bar{X}_n \in F_1) \\ &\le c_0[(r+1)/\delta_1]^{k(k-1)} \exp\{-n\delta\}. \end{aligned}$$

Hence, setting $\Delta = \varepsilon - \delta$, for some constants $c_1$, $c_2$ depending only on $k$ we have

$$\begin{aligned} &P_w(\phi_w(\bar{X}_n) > \varepsilon) \\ &\quad \le \{c_1(r+1)^{k(k-1)} + c_2[(r+1)/d]^{k(k-1)}\} \exp\{n\Delta\} \exp\{-n\varepsilon\}, \quad n \ge 1. \end{aligned} \tag{3.5}$$

We now estimate $r$, $d$ in terms of $\varepsilon$, $\Delta$. For $x \in bdA_{\varepsilon,w}$ we have

$$\varepsilon = \phi_w(x) = (x - \dot{c}(w))' \ddot{\phi}(\eta)(x - \dot{c}(w)),$$

where $\eta$ is a point on the line segment with endpoints $\dot{c}(w)$, $x$. Thus,

$$\begin{aligned}\varepsilon &\geq \|x - \dot{c}(w)\|^2 \inf\{\lambda_{\min}(z): z \in A_{\varepsilon,w}\}/2 \\ &= \|x - \dot{c}(w)\|^2/[2\sup\{\mu_{\max}(\omega): \omega \in I_{\varepsilon,w}\}] \\ &\geq \|x - \dot{c}(w)\|^2/[2\sup\{\operatorname{tr}\ddot{c}(\omega): \omega \in I_{\varepsilon,w}\}],\end{aligned}$$

where $\lambda_{\min}(z)$ is the minimum eigenvalue of $\ddot{\phi}(z)$, $\mu_{\max}(\omega)$ is the maximum eigenvalue of $\ddot{c}(\omega)$ and we used the facts that $\ddot{\phi}(\dot{c}(\omega)) = \ddot{c}^{-1}(\omega)$ for $\omega \in \Omega$, and $\dot{\phi}(A_{\varepsilon,w}) = I_{\varepsilon,w}$. Setting $s(\varepsilon) = \sup\{\operatorname{tr}\ddot{c}(\omega): \omega \in I_{\varepsilon,w}\}$, we then have $\|x - \dot{c}(w)\| \leq (2s(\varepsilon)\varepsilon)^{1/2}$. Since $x$ is an arbitrary boundary point of $A_{\varepsilon,w}$, letting $D$ denote the diameter of $A_{\varepsilon,w}$, it follows from the last inequality that $D \leq 2(2s(\varepsilon)\varepsilon)^{1/2}$. Finally, since $r \leq D$ (see Eggleston, 1969, Theorem 49), by Assumption 3.3, we obtain the following estimate of $r$:

$$r \leq 2(2f(\varepsilon)\varepsilon)^{1/2}. \tag{3.6}$$

We now estimate $d$. For $x \in bdA_{\delta,w}$, $y \in bdA_{\varepsilon,w}$ we have

$$\phi_w(y) = \phi_w(x) + (y - x)'\dot{\phi}_w(\eta), \tag{3.7}$$

where $\eta$ is a point on the line segment with endpoints $x$, $y$. Moreover, for $z \in A_{\varepsilon,w}$, $\dot{\phi}_w(z) = \ddot{\phi}(\zeta)(z - \dot{c}(w))$ for some $\zeta$ on the line segment with endpoints $\dot{c}(w)$, $z$. Hence,

$$\begin{aligned}\|\dot{\phi}_w(z)\| &\leq \|z - \dot{c}(w)\| \sup\{\lambda_{\max}(v): v \in A_{\varepsilon,w}\} \\ &= \|z - \dot{c}(w)\|/\inf\{\mu_{\min}(\omega): \omega \in I_{\varepsilon,w}\} \\ &\leq (2s(\varepsilon)\varepsilon)^{1/2}[s(\varepsilon)]^{k-1}/\inf\{\det\ddot{c}(\omega): \omega \in I_{\varepsilon,w}\} \\ &= (2s(\varepsilon)\varepsilon)^{1/2}[s(\varepsilon)]^{k-1}/i(\varepsilon),\end{aligned}$$

where $\lambda_{\max}(v)$ is the maximum eigenvalue of $\ddot{\phi}(v)$, $\mu_{\min}(\omega)$ is the minimum eigenvalue of $\ddot{c}(\omega)$, $i(\varepsilon) = \inf\{\det\ddot{c}(\omega): \omega \in I_{\varepsilon,w}\}$ and we used the fact that $\mu_{\min}(\omega) \geq \det\ddot{c}(\omega)/[\operatorname{tr}\ddot{c}(\omega)]_{k-1}$. Thus,

$$\sup\{\|\dot{\phi}_w(z)\|: z \in A_{\varepsilon,w}\} \leq (2s(\varepsilon)\varepsilon)^{1/2}[s(\varepsilon)]^{k-1}/i(\varepsilon).$$

From (3.7) we then obtain

$$\Delta = \varepsilon - \delta = (y - x)'\dot{\phi}_w(\eta) \leq \|y - x\| \sup\{\|\dot{\phi}_w(z)\|: z \in A_{\varepsilon,w}\},$$

or $\|y - x\| \geq i(\varepsilon)\Delta/\{(2s(\varepsilon)\varepsilon)^{1/2}[s(\varepsilon)]^{k-1}\}$. Therefore, by Assumption 3.3, we obtain the following estimate of $d$:

$$d \geq \gamma\Delta\exp\{-\beta\varepsilon\}/\{(2f(\varepsilon)\varepsilon)^{1/2}[f(\varepsilon)]^{k-1}\}. \tag{3.8}$$

By taking $\Delta = 1/n$ (i.e., $\delta = \varepsilon - 1/n$) and substituting the estimates of $r$ and $d$ ((3.6) and (3.8)) into (3.5) we obtain

$$P_w(\phi_w(\bar{X}_n) > \varepsilon) \leq [f_1(\varepsilon)\exp\{-k_0\varepsilon\} + n^{k(k-1)}f_2(\varepsilon)]\exp\{-(n - k_0)\varepsilon\}, \tag{3.9}$$

for all $n \geq 1$, $\varepsilon > 1/n$, where, in view of Assumption 3.3, $f_1$ and $f_2$ are (positive-valued) Borel-measurable functions depending in general on $w$, bounded on any finite interval $(0, \delta)$ and such that $f_i(t) = \exp\{o(t)\}$ as $t \to \infty$, $i = 1, 2$. Note that these properties of $f_i$ imply

$$\int_0^\infty f_i(t)\exp\{-at\}\,dt < \infty \qquad \text{for all} \quad a > 0, \quad \text{and}$$
$$(3.10)\qquad \int_0^\infty f_i(t)\exp\{-at\}\,dt = O\left(\frac{1}{a}\right) \quad \text{as} \quad a \to \infty, \quad i = 1, 2.$$

Consider now $n > \max(m - 1, k_0)$ and $0 < \tau < 1$. Then, by Assumption 3.2, $\exp\{\tau(n - k_0)\phi_w(\bar{X}_n)\}$ is a bona fide random variable under $P_w$. Using the bound in (3.9) we obtain

$$\begin{aligned}
&E_w\exp\{\tau(n - k_0)\phi_w(\bar{X}_n)\}\\
&\quad = \int_0^\infty P_w(\exp\{\tau(n - k_0)\phi_w(\bar{X}_n)\} > t)\,dt\\
&\quad \le e + \int_e^\infty P_w\left(\phi_w(\bar{X}_n) > \frac{\log t}{\tau(n - k_0)}\right) dt\\
&\quad = e + \tau(n - k_0)\int_{1/[\tau(n-k_0)]}^\infty P_w(\phi_w(\bar{X}_n) > t)\exp\{\tau(n - k_0)t\}\,dt\\
&\quad \le e + \tau(n - k_0)\int_{1/[\tau(n-k_0)]}^\infty f_1(t)\exp\{-(1 - \tau)nt - \tau k_0 t\}\,dt\\
&\qquad + \tau n^{k(k-1)}(n - k_0)\int_{1/[\tau(n-k_0)]}^\infty f_2(t)\exp\{-(1 - \tau)(n - k_0)t\}\,dt.
\end{aligned}$$

Hence, by (3.10), the first two assertions of the theorem now follow.

The last one simply follows from (3.3) and Markov's inequality. □

REMARK 1. The multinomial distribution. An essential requirement in our method of inscribing a polytope between two level sets of $\phi_w$ is that the boundaries of these sets do not touch each other. See Theorem 3.2 and Proposition 5.6 of the Appendix. In view of the continuity of $\phi_w$, a sufficient condition for achieving this is to assume that $B$ is open, as we did. An important model for which $B$ is closed and hence Theorem 3.2 cannot be invoked is the multinomial distribution with all the cell probabilities positive (to fall into the exponential framework). This model, however, can be analyzed by a direct calculation. To this end, let $X_1, X_2, \cdots$ be a sequence of i.i.d. random vectors in $R^k$, $k \ge 2$, following the multinomial distribution, and $\bar{X}_n = (\bar{X}_{n1}, \cdots, \bar{X}_{nk})$ be the sample mean of $X_1, \cdots, X_n$, $n \ge 1$. The sample space is

$$S = \{(1, 0, \cdots, 0)', \cdots, (0, 0, \cdots, 0, 1)'\} \subset R^k$$

and the parameter space is

$$\Pi = \{\pi = (\pi_1, \cdots, \pi_k)' \in R^k\colon \pi_i \ge 0, \textstyle\sum_{i=1}^k \pi_i = 1\}.$$

Consider next $\eta = (\eta_1, \cdots, \eta_k)' \in \Pi$ and let $\ell_n$ be the likelihood ratio statistic for testing $H_0$: $\pi = \eta$ vs. $H_1$: $\pi \neq \eta$. Then,

$$\begin{aligned} E_\eta \ell_n &= E_\eta[\sup\{\pi_1^{n\bar{X}_{n1}} \cdots \pi_k^{n\bar{X}_{nk}}: \pi \in \Pi\}/(\eta_1^{n\bar{X}_{n1}} \cdots \eta_k^{n\bar{X}_{nk}})] \\ &= E_\eta[\bar{X}_{n1}^{n\bar{X}_{n1}} \cdots \bar{X}_{nk}^{n\bar{X}_{nk}}/(\eta_1^{n\bar{X}_{n1}} \cdots \eta_k^{n\bar{X}_{nk}})] \\ &= \textstyle\sum_{s_1+\cdots+s_k=n, s_i \text{integers} \geq 0} [(s_1/n)^{s_1} \cdots (s_k/n)^{s_k}/(\eta_1^{s_1} \cdots \eta_k^{s_k})] \\ &\quad \cdot [n!/(s_1! \cdots s_k!)]\eta_1^{s_1} \cdots \eta_k^{s_k} \\ &= \sum [n!/(s_1! \cdots s_k!)](s_1/n)^{s_1} \cdots (s_k/n)^{s_k} \leq \binom{n+k-1}{k-1}, \end{aligned}$$

since each summand is a multinomial probability and the number of integral solutions of the equation $s_1 + \cdots + s_k = n$ is $\binom{n+k-1}{k-1}$. Hence, for some constant $\alpha(k)$ depending only on $k$ we obtain $E_\eta \ell_n \leq \alpha(k)n^{k-1}$. Letting now $\phi_\eta = \log \ell_n/n$ we conclude that $P_\eta(\phi_\eta(\bar{X}_n) > \varepsilon) \leq \alpha(k)n^{k-1}\exp\{-n\varepsilon\}$, for all $\varepsilon > 0$, $n \geq 1$, $\eta \in \Pi$. Note that the same result is obtained by applying inequality (2.8) in Hoeffding (1965). See also Kallenberg (1978, page 68).

REMARK 2. The case of independent components. When the data vectors $X_1, X_2, \cdots$ (distributed according to an exponential family) consist of independent components it is possible to obtain results similar to (3.3) and (3.4) (in fact, sharper) by simply applying Lemma 1.1 to each of the components without having to verify Assumptions 3.1–3.3. The details are as follows.

We introduce some notation we will need below. Let $\nu_i$, $i = 1, \cdots, k$, $k \geq 1$, be nondegenerate $\sigma$-finite measures on $\mathscr{B}(R)$ and

$$\Omega_i = \left\{\omega \in R: \exp\{c_i(\omega)\} = \int \exp\{\omega y\}\, d\nu_i(y) < \infty\right\}$$

be nonempty and open. Let $\nu = \nu_1 \times \cdots \times \nu_k$ be the product measure of $\nu_i$'s, and $\mathscr{P} = \{P_\omega: \omega \in \Omega\}$ the natural exponential family generated by $\nu$. Clearly, the natural parameter space for $\mathscr{P}$ is the open set $\Omega = \Omega_1 \times \cdots \times \Omega_k$ and the corresponding $c(\omega) = \sum_{i=1}^k c_i(\omega_i)$, $\omega = (\omega_1, \cdots, \omega_k)' \in \Omega$. If $X = (Y_1, \cdots, Y_k)' \sim P_\omega$, $Y_i$, $i = 1, \cdots, k$ are independent with densities $\exp\{\omega_i y - c_i(\omega_i)\}$ with respect to $\nu_i$. Furthermore, letting, for $\partial_i \in \Omega_i$, $\phi_i(y \mid \partial_i) = \sup\{(\omega_i - \partial_i)y - c_i(\omega_i) + c_i(\partial_i): \omega_i \in \Omega_i\}$, we have

$$\phi_\partial(x) = \sup\{(\omega - \partial)'x - c(\omega) + c(\partial): \omega \in \Omega\} = \textstyle\sum_{i=1}^k \phi_i(y_i \mid \partial_i),$$

$x = (y_1, \cdots, y_k)'$, $\partial = (\partial_1, \cdots, \partial_k)'$. Let $X_i = (X_{i1}, \cdots, X_{ik})'$, $i = 1, 2, \cdots$ be a sequence of i.i.d. random vectors with family of distributions $\mathscr{P}$ and $\bar{X}_n = (\bar{X}_{n1}, \cdots, \bar{X}_{nk})'$, where $\bar{X}_{nj} = \sum_{i=1}^n X_{ij}/n$, $j = 1, \cdots, k$, $n \geq 1$. Then, given $\varepsilon > 0$, $0 < \tau < 1$,

$$\begin{aligned} P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon) &\leq \exp\{-\tau n\varepsilon\}E_\partial \exp\{\tau n \phi_\partial(\bar{X}_n)\} \\ &= \exp\{-\tau n\varepsilon\} \textstyle\prod_{i=1}^k E_{\partial_i}\exp\{\tau n \phi_i(\bar{X}_{ni} \mid \partial_i)\}, \end{aligned}$$

and hence

$$P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon) \le [1 + 2\tau/(1-\tau)]^k \exp\{-\tau n\varepsilon\}, \quad n \ge 1, \tag{3.11}$$

where we used the fact that $E_{\partial_i} \exp\{\tau n \phi_i(\bar{X}_{ni} \mid \partial_i)\} \le 1 + 2\tau/(1-\tau)$, $i = 1, \cdots, k$, which follows immediately from Lemma 1.1 by integration. An improvement on the above bound can be obtained by minimizing the right-hand side expression in (3.11) with respect to $\tau \in (0, 1)$, the minimizing choice being $\tau_0 = [1 - 2k/(n\varepsilon)]^{1/2}$, $\varepsilon > 2k/n$. Of interest is also the choice $\tau_1 = 1 - 1/(n\varepsilon)$ giving

$$P_\partial(\phi_\partial(\bar{X}_n) > \varepsilon) \le e(2n\varepsilon - 1)^k \exp\{-n\varepsilon\}, \quad n \ge 1, \quad \varepsilon > 1/n. \tag{3.12}$$

REMARK 3. For a bound on $P_w(\phi_w(\bar{X}_n) > \varepsilon)$ that does not involve $\tau$ one can take $\tau = 1 - 1/(n - k_0)$ and argue as in the closing lines of the proof of Theorem 3.2 to obtain

$$E_w \exp\{(n - k_0 - 1)\phi_w(\bar{X}_n)\} = O(n^{k(k-1)+1})$$

and hence, for some constant $\kappa(w)$,

$$P_w(\phi_w(\bar{X}_n) > \varepsilon) \le \kappa(w) n^{k(k-1)+1} \exp\{-(n - k_0 - 1)\varepsilon\}, \tag{3.13}$$

for all $\varepsilon > 0$, $n > \max(m - 1, k_0 + 1)$.

REMARK 4. When $f$ in Assumption 3.3 is a polynomial (with necessarily positive constant term since $f(\varepsilon) \ge \operatorname{tr} \ddot{c}(w)$ for all $\varepsilon > 0$)

$$\int_0^\infty f_2(t) \exp\{-at\}\, dt = O(1/a^{k(k-1)/2+1})$$

as $a \to \infty$. This follows immediately by displaying $f_2$ (which is feasible from (3.5), (3.6), (3.8), and (3.9)) and integrating out. Hence, arguing as in the proof of Theorem 3.2, we obtain, for $0 < \tau < 1$,

$$E_w \exp\{\tau(n - k_0)\phi_w(\bar{X}_n)\} = O(n^{k(k-1)/2}), \tag{3.14}$$

which provides a significant improvement on (3.3). This situation occurs in the examples of Section 4.

As an immediate consequence of Theorem 3.2 (more precisely, of (3.13)) we generalize below results in Hoeffding (1965) for the multinomial distribution and in Herr (1967) for the multivariate normal distribution, and thus give an answer to a question raised by Chernoff (see discussion following Hoeffding's paper). For $\partial \in \Omega$, $\Omega_1 \subset \Omega$ we let $K(\Omega_1, \partial) = \inf\{K(\omega, \partial): \omega \in \Omega_1\}$.

THEOREM 3.3. *Suppose that Assumptions* 3.1–3.3 *hold, and let* $\hat{\omega}_n$ *be the maximum likelihood estimator of* $\omega \in \Omega$ *and* $\Omega_1$ *be a Borel-measurable subset of* $\Omega$. *Then, for some constant* $\kappa(w)$,

$$P_w(\hat{\omega}_n \in \Omega_1) \le \kappa(w) n^{k(k-1)+1} \exp\{-(n - k_0 - 1)K(\Omega_1, w)\}, \tag{3.15}$$

*for all* $n > \max(m - 1, k_0 + 1)$ *uniformly in* $\Omega_1$.

**Proof.** To avoid trivialities, assume $K(\Omega_1, w) > 0$. Then, for

$$n > \max(m-1, k_0+1)$$

we have

$$\begin{aligned} P_w(\hat{\omega}_n \in \Omega_1) &= P_w(\dot{\phi}(\bar{X}_n) \in \Omega_1) \le P_w(K(\dot{\phi}(\bar{X}_n), w) \ge K(\Omega_1, w)) \\ &= P_w(\phi_w(\bar{X}_n) \ge K(\Omega_1, w)) \le \kappa(w) n^{k(k-1)+1} \exp\{-(n-k_0-1)K(\Omega_1, w)\}, \end{aligned}$$

by (3.13). □

**Remark 5.** It turns out that (3.15) is not just a consequence of (3.13), but in fact equivalent to it. This is straightforward to show.

For one-dimensional exponential families an analogous result can be obtained by applying Lemma 1.1.

## 4. Examples.

In this section we present two examples for which Assumptions 3.1–3.3 are satisfied. Detailed arguments as well as other examples are contained in Kourouklis (1981).

**Example 1.** The family of gamma distributions with densities $g(y \mid \omega_1, \omega_2) = \omega_1^{\omega_2} y^{\omega_2} \exp\{-\omega_1 y\}/\Gamma(\omega_2)$, $y, \omega_1, \omega_2 > 0$, with respect to the measure $dy/y$. Assumption 3.3 is satisfied for $w = (1, 1)'$.

**Example 2.** The family of $p$-variate normal distributions $N_p(\mu, \Sigma)$, $(\mu, \Sigma) \in \Theta = \{(\mu, \Sigma): \mu \in R^p, \Sigma \text{ is } p \times p \text{ positive definite}\}$, $p \ge 1$. Let $\mathscr{P} = \{P_\omega: \omega \in \Omega\}$ denote the corresponding $k$-dimensional natural exponential family, where $k = 2p + p(p-1)/2$, and $a$ denote the transformation that maps $\Theta$ $1-1$ onto the natural parameter space $\Omega$. Also let $X_1, X_2, \cdots$ be a sequence of i.i.d. random vectors in $R^k$ with distribution $P_\omega$, $\omega \in \Omega$. Since (as is well known) $\bar{X}_n = \sum_{i=1}^n X_i/n$ has an absolutely continuous distribution for $n \ge p+1$, Assumptions 3.1 and 3.2 are satisfied with $m = p+1$. We will show next that Assumption 3.3 holds with $w = a[(0, I_p)]$, where $I_p$ is the $p \times p$ identity matrix. Note first that here

$$I_{\varepsilon,w} = \{\omega = a[(\mu, \Sigma)] \in \Omega: \mu'\mu + \operatorname{tr} \Sigma - \log \det \Sigma - p \le 2\varepsilon\}, \quad \varepsilon > 0.$$

For $\omega = a[(\mu, \Sigma)] \in \Omega$, direct calculations yield the crude bound

$$\operatorname{tr} \ddot{c}(\omega) \le \operatorname{tr} \Sigma(1 + 2\mu'\mu + 3 \operatorname{tr} \Sigma/2). \tag{4.1}$$

Direct calculations and induction on $p$ also yield

$$\det \ddot{c}(\omega) = 2^{-p}(\det \Sigma)^{p+2}. \tag{4.2}$$

Consider now $\omega = a[(\mu, \Sigma)] \in I_{\varepsilon,w}$ and let $\lambda_1, \cdots, \lambda_p$ be the eigenvalues of $\Sigma$. Then $\mu'\mu + \sum_{i=1}^p (\lambda_i - \log \lambda_i - 1) \le 2\varepsilon$, which in turn entails

$$\mu'\mu \le 2\varepsilon \quad \text{and} \quad \lambda_i - \log \lambda_i \le 2\varepsilon + 1, \quad i = 1, \cdots, p, \tag{4.3}$$

since $x - \log x - 1 \geq 0$ for $x > 0$. The last set of inequalities implies

$$\exp\{-2\varepsilon - 1\} \leq \lambda_i \leq (2\varepsilon + 1)/(1 - e^{-1}), \quad i = 1, \cdots, p. \tag{4.4}$$

(For the right-hand side inequality, use the fact $x \log x \geq -e^{-1}$, for $x > 0$.) From (4.1)–(4.4) it clearly follows that Assumption 3.3 holds with $\beta = 2p(p + 2)$, $\gamma = 2^{-p}\exp\{-p(p + 2)\}$, and

$$f(\varepsilon) = p(2\varepsilon + 1)/(1 - e^{-1}) + 2(p + 1)\varepsilon(2\varepsilon + 1)/(1 - e^{-1}) + 3p^2(2\varepsilon + 1)^2/[2(1 - e^{-1})^2].$$

We should note here that for the multivariate normal distribution, inequality (2.7) in Herr (1967) states

$$P_\omega(\phi_\omega(\bar{X}_n) > \varepsilon) \leq An^{p(p+1)/2}\exp\{-(n - 2p - 3)\varepsilon\}, \tag{4.5}$$

for all large $n$ uniformly in $\varepsilon > 0$ and $\omega \in \Omega$, where $A$ is a constant, and hence implies the (general) result (3.4) (as well as (3.13) and (3.14)). However, the above bound can also be improved. Indeed, by direct computation,

$$E_\omega \exp\{(n - p - 1)\phi_\omega(\bar{X}_n)\} = O(n^{p+p(p-1)/4})$$

uniformly in $\omega \in \Omega$ and hence

$$P_\omega(\phi_\omega(\bar{X}_n) > \varepsilon) \leq A_1 n^{p+p(p-1)/4}\exp\{-(n - p - 1)\varepsilon\}, \tag{4.6}$$

for all $n > p + 1$, $\varepsilon > 0$, $\omega \in \Omega$, where $A_1$ is a constant depending only on $p$.

Lastly note that direct computation also yields

$$E_\omega \exp\{\tau(n - p)\phi_\omega(\bar{X}_n)\} = O(1) \tag{4.7}$$

for any $0 < \tau < 1$, uniformly in $\omega \in \Omega$ (Kourouklis, 1981). This is therefore the best order of magnitude one could hope to get in (3.3).

**5. Appendix.** Using results of Bronshteyn and Ivanov (1975) and Dudley (1974), it is shown below (in Proposition 5.6) how a polytope can be inscribed between two bounded convex sets in $R^k$, $k \geq 2$, one contained in the other. A crude estimate of the number of "sides" of this polytope is obtained. Preliminary definitions and lemmas are in order.

DEFINITION 5.1. A $\delta$-grid for a set $B \subset R^k$ is a subset $A$ of $B$, such that each point of $B$ is within (Euclidean) distance $\delta$ from some point of $A$.

LEMMA 5.2. *Given a sphere $S$ of radius $r > 0$ in $R^k$ and $0 < \delta < 4r$, there is a $\delta$-grid for the boundary of $S$, bdS, which contains at most $c(r/\delta)^{k-1}$ points, where $c$ is a positive constant depending only on $k$.*

PROOF. The proof is essentially contained in the proof of Lemma 1 of Bronshteyn and Ivanov (1975). See also Appendix A of Kourouklis (1981). □

LEMMA 5.3 (Bronshteyn and Ivanov, 1975). *Let $C$ be a bounded convex set in*

*$R^k$ and $x_1, \cdots, x_m$, $m \geq 1$, points in $R^k \sim \mathrm{cl}C$ (cl$C$ denotes the closure of $C$). Suppose that for every supporting hyperplane of* cl$C$ *there is at least one point $x_i$ lying in the open half-space disjoint from the half-space containing* cl$C$. *Then the convex hull of $\{x_1, \cdots, x_m\}$ contains $C$.*

DEFINITION 5.4 (Dudley, 1974). A set $C$ in $R^k$ is called analytic if there is an entire analytic function $g$ such that $C = \{x \in R^k: g(x) \leq 1\}$ and the gradient of $g$, $\dot{g}$, is nonzero on $bdC$.

LEMMA 5.5 (Bronshteyn and Ivanov, 1975; Dudley, 1974). *Let $x, y, u, v$ be points in $R^k$ such that $(x - y)'u \geq 0$ and $(x - y)'v \leq 0$. Then $\| x + u - y - v \| \geq \max\{\| x - y \|, \| u - v \|\}$.*

PROPOSITION 5.6. *Let $C_1 \subset C_2$ be sets in $R^k$ such that $C_1$ is convex and analytic, $C_2$ is bounded and convex, and $d = \inf\{\| x_1 - x_2 \|: x_1 \in bdC_1, x_2 \in bdC_2\} > 0$. Then there is a polytope $P$ satisfying $C_1 \subset P \subset C_2$ and such that it is the intersection of at most $c_0[(r + 1)/\delta]^{k(k-1)}$ closed half-spaces, where $r$ is the circumradius of $C_2$, $\delta = \min(\pi/8, d/2)$ and $c_0$ a positive constant depending only on $k$.*

PROOF. The proof is along the lines of Lemmas 4.4 and 4.5 of Dudley (1974). Let $S$ be the sphere of radius $r + 1$ co-centric to the circumsphere of $C_2$. By Lemma 5.2, there is a $\delta$-grid, say $S_\delta$, for $bdS$ which contains at most $c[(r + 1)/\delta]^{k-1}$ points, $c$ being as in Lemma 5.2. For every $p \in bdS$ there is a unique nearest point $n(p) \in bdC_1$, with $\| p - n(p) \| \geq 1$. The function $n(\cdot)$ maps $bdS$ $1 - 1$ onto $bdC_1$. Note that $p - n(p)$ is normal of cl$C_1$ at $n(p)$ (a point $x_0$ is normal of a convex set $C$ at $a \in C$ if $(x - a)'x_0 \leq 0$ for all $x \in C$). It is next shown that $A = \{n(p): p \in S_\delta\}$ is a $\delta$-grid for $bdC_1$. Let $n(q)$, $q \in bdS$, be a point in $bdC_1$. There is a point $p \in S_\delta$ such that $\| p - q \| < \delta$. Let $u = p - n(p)$, $v = q - n(q)$. Since $u, v$ are normal of $clC_1$ at $n(p)$, $n(q)$ respectively, we have $(n(p) - n(q))'u \geq 0$ and $(n(p) - n(q))'v \leq 0$. Apply now Lemma 5.5 to obtain that $\| n(p) - n(q) \| < \delta$ and $\| u - v \| < \delta$. Consider next the set

$$D = \{\gamma(p - n(p))/\| p - n(p) \|: p \in S_\delta\},$$

where $\gamma = 2\delta$. We will show that the assumptions of Lemma 5.3 hold for $C_1$ and the points of $D$. Consider a hyperplane, say $H$, which supports cl$C_1$ at $n(q) \in bdC_1$, $q \in bdS$. There is $p \in bdS$ such that $\| n(p) - n(q) \| < \delta$. Let $u, v$ be as above and $u_1 = u/\| u \|$, $v_1 = v/\| v \|$. Since $\| u \| \geq 1$, $\| v \| \geq 1$ we have $\| u_1 - v_1 \| \leq \| u - v \| < \delta$. Let $\theta \in [0, \pi]$ be (the smallest nonnegative) angle between $u_1$, $v_1$. Since $\| u_1 - v_1 \| = 2 \sin(\theta/2)$ and $\theta \leq \pi \sin(\theta/2)$ for $0 \leq \theta \leq \pi$ by concavity, we have $\theta \leq \pi\delta/2 < 2\delta \leq \pi/4$. Let $\alpha$ be the distance of $n(p)$ to $H$ along the normal $p - n(p)$. Since $0 \leq \theta \leq \pi/4$, $\alpha \leq \| n(p) - n(q) \| \tan \theta < \delta$. Hence, $\gamma > \alpha$ and $\gamma(p - n(p))/\| p - n(p) \|$ lies in the open half-space determined by $H$ disjoint from the half-space containing cl$C_1$. Apply now Lemma 5.3 to conclude that $C_1 \subset$ conv $D$ (convex hull of $D$). Since $\gamma \leq d$, $D \subset C_2$, hence conv $D \subset C_2$. By taking $P =$ conv $D$, we have $C_1 \subset P \subset C_2$. Furthermore, we can write $P = \cap \{H_j: 1 \leq j \leq f_{k-1}(P)\}$, where $f_{k-1}(P)$ is the number of $(k - 1)$-dimensional faces of $P$ and $H_j$ the closed half-space associated with these faces and containing $P$ (see Grünbaum, 1967, pages 17, 31, 32). Since $f_{k-1}(P) \le \binom{f_0(P)}{k}$ (Grünbaum, 1967, page 31), where $f_0(P)$ is the number of exposed points of $P$ and the latter is less than or equal to the cardinality of $D$, we conclude that $f_{k-1}(P) \le c_0[(r+1)/\delta]^{k(k-1)}$, where $c_0 = c^k/k!$ □

REMARK 5.7. The above estimate $c_0[(r+1)/\delta]^{k(k-1)}$ of $f_{k-1}(P)$ is very crude. For $k = 2$ it can be improved to $c(r+1)/\delta$.

**Acknowledgements.** This paper forms a revised portion of the author's Ph.D. dissertation at Rutgers University. The author wishes to express his gratitude to Prof. R. H. Berk for his advice and support throughout the development of this work. Thanks are extended to Prof. M. Perlis of the Hebrew University for bringing to the author's attention the Bronshteyn and Ivanov (1975) and the Dudley (1974) references, and to Prof. W. C. M. Kallenberg for providing detailed explanations on a work of his. Stimulating comments and detailed suggestions for improvement by a referee and an Associate Editor are sincerely acknowledged.

## REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families.* Wiley, New York.

BERK, R. H. (1972). Consistency and asymptotic normality of mle's for exponential models. *Ann. Math. Statist.* **43** 193–204.

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

BOROVKOV, A. A. (1968). Some inequalities for sums of multidimensional random variables. *Theory Probab. Appl.* **13** 156–160.

BRONSHTEYN, E. M. and IVANOV, L. D. (1975). The approximation of convex sets by polyhedra. *Siberian Math. J.* **16** 852–853.

DUDLEY, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approx. Theory* **10** 227–236.

EGGLESTON, H. G. (1969). *Convexity.* Cambridge Univ. Press, Cambridge, England.

EFRON, B. and TRUAX, D. (1968). Large deviation theory in exponential families. *Ann. Math. Statist.* **39** 1403–1424.

GRÜNBAUM, B. (1967). *Convex Polytopes.* Interscience Publishers, New York.

HERR, D. G. (1967). Asymptotically optimal tests for multivariate normal distributions. *Ann. Math. Statist.* **38** 1829–1844.

HOEFFDING, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408.

KALLENBERG, W. C. M. (1978). *Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families.* Mathematical Centre Tracts, Amsterdam.

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

KOUROUKLIS, S. (1984). Bahadur optimality of sequential experiments for exponential families. *Ann. Statist.* **12** 1522–1527.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton Univ. Press, Princeton.

WOODROOFE, M. (1978). Large deviations of likelihood ratio statistics with applications to sequential testing. *Ann. Statist.* **6** 72–84.

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