The Annals of Statistics
1996, Vol. 24, No. 5, 2280–2287

A LARGE DEVIATION THEOREM FOR THE q-SAMPLE LIKELIHOOD RATIO STATISTIC

BY FRANTIŠEK RUBLÍK

Institute of Measurement of SAS, Bratislava

An upper bound for the tail probability $P_\theta(\log(L(x_{(n_1,\dots,n_q)},\Theta)/L(x_{(n_1,\dots,n_q)},\theta))\geq t)$ is derived in the case of sampling from q populations. This estimate is used for establishing the Hodges–Lehmann optimality of a test statistic for a hypothesis on exponential distributions.

1. Introduction and the main result. Let probabilities $\{\overline{P}_\gamma;\ \gamma\in\Xi\}$ be defined on $(X,\mathscr{F})$ by means of the densities

$$f(x,\gamma)=\frac{d\overline{P}_\gamma}{d\nu}(x). \tag{1.1}$$

The parameter space of overall parameters is the q-fold Cartesian product

$$\Theta=\Xi^q, \tag{1.2}$$

where in $\theta=(\theta_1,\dots,\theta_q)\in\Theta$ the symbol θ_j stands for the parameter of the jth population. The outcome of the sampling from the jth population will be denoted by $x(j,n_j)=(x_1^{(j)},\dots,x_{n_j}^{(j)})$. Thus

$$x_{(n_1,\dots,n_q)}=(x(1,n_1),\dots,x(q,n_q)) \tag{1.3}$$

is the pooled sample and its distribution is the product measure $P_\theta=\overline{P}_{\theta_1}^{n_1}\times\cdots\times\overline{P}_{\theta_q}^{n_q}$.

For $\Omega\subset\Theta$ let

$$L(x_{(n_1,\dots,n_q)},\Omega)=\sup\left\{\prod_{j=1}^{q}\prod_{i=1}^{n_j}f(x_i^{(j)},\theta_j);\ (\theta_1,\dots,\theta_q)\in\Omega\right\}. \tag{1.4}$$

An upper bound for the probability $P_\theta(\log(L(x_{(n_1,\dots,n_q)},\Theta)/L(x_{(n_1,\dots,n_q)},\theta))\geq t)$ can be useful for establishing asymptotic properties of the likelihood ratio test statistics under validity of the alternative hypothesis. Such upper estimates were derived in Theorem 3.2 in Kourouklis (1984) and can also be found in Kourouklis (1987). An essential assumption in these papers is that $f(x,\gamma)=\exp(\gamma'x-C(\gamma))$, that is, that the densities (1.1) form an exponential family of distributions. The aim of this paper is to present a theorem not requiring this property. Moreover, the imposed regularity conditions also admit densities discontinuous in θ.

Received April 1993; revised June 1996.
AMS 1991 *subject classifications*. Primary 60F10, 62F05; secondary 62E15, 62F12.
Key words and phrases. Large deviations, exponential distributions with unknown lower bound, Hodges–Lehmann optimality.

(C1). Ξ is a subset of R^m and the function (1.4) is measurable whenever $\Omega = \Theta \cap C$, where $C \subset R^{mq}$ is either open or closed.

(C2). Let ρ denote the usual Euclidean distance and

$$V(\gamma^*, \delta^*) = \{\tilde{\gamma} \in \Xi;\ \rho(\gamma^*, \tilde{\gamma}) < \delta^*\}, \qquad L(x, V) = \sup\{f(x, \tilde{\gamma});\ \tilde{\gamma} \in V\}. \tag{1.5}$$

For every $\gamma \in \Xi$ there exist numbers $M = M(\gamma)$, $\delta = \delta(\gamma)$ and $\varepsilon = \varepsilon(\gamma)$ such that the inequality

$$\int L(x, V(\gamma^*, \delta^*))\, d\nu(x) \le 1 + M\delta^* \tag{1.6}$$

holds for each $\gamma^* \in V(\gamma, \delta)$ and $0 < \delta^* \le \varepsilon$.

(C3). Let the total and the relative sample sizes defined by the formulas

$$n = n_1 + \cdots + n_q, \qquad \hat{p}_j = \frac{n_j}{n} \tag{1.7}$$

be such that

$$n \to \infty, \qquad \hat{p}_j \to p_j > 0, \qquad j = 1, \ldots, q. \tag{1.8}$$

Let $\theta \in \Theta$. For every $a \in (0, +\infty)$ there exists a compact subset Γ of Θ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_\theta\big(L(x_{(n_1,\ldots,n_q)}, \Theta - \Gamma) \ge L(x_{(n_1,\ldots,n_q)}, \theta)\big) < -a. \tag{1.9}$$

We remark that for the sake of simplicity we omit in the notation in (C3) an index for the experiment. Thus we tacitly assume that $n_j = n_j^{(u)}$, $n = n^{(u)}$ denote sample sizes in the uth experiment, $u = 1, 2, \ldots,$ and the limits in (C3) are related to u tending to infinity.

THEOREM 1.1. *Suppose that the conditions* (C1) *and* (C2) *hold.*

(i) *If Γ is a nonempty compact subset of Θ, then there exists a constant $C = C(\Gamma)$ such that*

$$P_\theta\left(\log \frac{L(x_{(n_1,\ldots,n_q)}, \Gamma)}{L(x_{(n_1,\ldots,n_q)}, \theta)} \ge nt\right) \le \exp\left(-nt + m \sum_{j=1}^{q} \log n_j + C\right) \tag{1.10}$$

for all $\theta \in \Theta$, all sample sizes $n_1, \ldots, n_q$ and all real numbers t.

(ii) *Furthermore, suppose that $\theta = (\theta_1, \ldots, \theta_q)$ belongs to Θ, the relations* (1.8) *hold and the condition* (C3) *is satisfied. If A is a positive real number, then there exists a constant $C = C(\theta, A, \{n_1\}, \ldots, \{n_q\})$ such that*

$$P_\theta\left(\log \frac{L(x_{(n_1,\ldots,n_q)}, \Theta)}{L(x_{(n_1,\ldots,n_q)}, \theta)} \ge nt\right) \le \exp\left(-nt + m \sum_{j=1}^{q} \log n_j + C\right) \tag{1.11}$$

for all $t \le A$ and all sample sizes occurring in (1.7).

An upper bound analogous to (1.11) and valid for all $t \in (n^{-1}, A)$ is presented in Lemma 4.4 of Kourouklis (1987), where the underlying class of probabilities is required to form an exponential family of distributions for which the maximum likelihood estimate exists with probability 1 for n sufficiently large. In the one-sample setting these exponential families are treated in Kourouklis (1984), where instead of m the one-sample formulas (3.4) and (3.13) contain the terms $m(m-1)$ and $m(m-1)+1$, respectively, and the resulting inaccuracy is worse than that in (1.11), but on the other hand they yield bounds uniform in $t \geq 0$. The mentioned exponential families also include k-dimensional nonsingular normal distributions. A multisample version of the upper bound uniform in $t \geq 0$ for these normal distributions is presented in Lemma 3.1 of Rublík (1995), where the obtained inaccuracy is asymptotically a little better than that in the one-sample formula (4.6) of Kourouklis (1984). For the Poisson distributions a uniform large deviation upper bound has been established in the multisample setting in Lemma 4.3 in Kourouklis (1987). Since multinomial distributions can be represented with a compact parameter space, from Theorem 1.1 we obtain that (1.10) holds also for this class of distributions [the same inequality but with a term $C = C(n_1, \dots, n_q)$ better than that given by (2.9) can be proved in this case similarly to Theorem 2.1 in Hoeffding (1965)]. As examples of nonexponential classes of distributions satisfying the assumptions of Theorem 1.1(ii), we mention here the Laplace distributions and the exponential distributions with unknown lower bound.

The reasoning used in proving (1.10) is of the type related to Bahadur (1965) and utilizes the structure of the Euclidean space. The inaccuracy order $\mathscr{O}(\log n)$, which follows from the inequality (1.11), is better than the order obtained in the one-sample case in Lemma 5 in Bahadur (1965), which on the other hand yields an upper bound uniform in t.

Theorem 1.1 is the main result of the paper and in this general setting can be used for proving asymptotic optimality of the likelihood ratio test statistics for testing simple hypotheses. A formula on large deviation probabilities derived by means of application of the theorem to the exponential distributions is denoted by (3.2) in Section 3 and is used for proving asymptotic optimality of a test of a composite hypothesis.

As to the assumptions of Theorem 1.1, we remark that the multisample conditions (C1) and (C3) can be imposed also in terms of one-sample properties of the densities (1.1). It is shown in the proof of Lemma 2.4 in Rublík (1989) that the condition (C1) is fulfilled if the following assumptions are satisfied: Ξ is a σ-compact subset of R^m, $f(x, \cdot)$ is an upper semicontinuous function for each $x \in X$ and there exists a countable subset $\tilde{\Xi}$ of Ξ containing for every $\gamma \in \Xi$ a sequence $\{\gamma_k\}_{k=1}^{\infty}$ such that $\gamma_k \to \gamma$ and $f(x, \gamma_k) \to f(x, \gamma)$ for each $x \in X$. Further, according to Lemma 2.2(iii) in Rublík (1989), the validity of (C1) and the following condition imply (C3).

(D1). For every $\gamma \in \Xi$ and every pair of positive real numbers a, c there exist measurable sets $A_n \subset X^n$, a nonempty compact set $\Gamma \subset \Xi$, an integer N

and a constant $\varepsilon > 0$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_\gamma(A_n) < -a, \tag{1.12}$$

$$\inf\left\{\frac{1}{n}\log L(x_1,\dots,x_n,\gamma);\ (x_1,\dots,x_n)\in X^n - A_n,\ n\ge N\right\} > -\infty, \tag{1.13}$$

$$\sup\left\{\frac{1}{n}\log L(x_1,\dots,x_n,\Xi-\Gamma);\ (x_1,\dots,x_n)\in X^n - A_n,\ n\ge N\right\} < -c, \tag{1.14}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log P_\gamma\left(\frac{1}{n}\log\frac{L(x_1,\dots,x_n,\Xi)}{L(x_1,\dots,x_n,\gamma)} \ge \varepsilon\right) < -a. \tag{1.15}$$

Although this condition looks complicated, in some cases it is easier to verify than (C3).

2. Proof of Theorem 1.1. (i) Since the projection of Γ onto the jth coordinate space is also a compact set, we see that $\Gamma \subset \Gamma_1 \times \cdots \times \Gamma_q$, where $\Gamma_1,\dots,\Gamma_q$ are compact subsets of Ξ. Hence

$$P_\theta\left(\log\frac{L(x_{(n_1,\dots,n_q)},\Gamma)}{L(x_{(n_1,\dots,n_q)},\theta)} \ge nt\right) \le e^{-nt}\prod_{j=1}^{q}\int L(x(j,n_j),\Gamma_j)\,d\nu(x(j,n_j)), \tag{2.1}$$

where $d\nu(x_1^{(j)},\dots,x_{n_j}^{(j)})$ denotes integration with respect to the n_j-fold product measure $\nu\times\cdots\times\nu$.

Let $j\in\{1,\dots,q\}$ be an arbitrary fixed index. Since the set Γ_j is compact, there exist finitely many points $\gamma_1,\dots,\gamma_v$ such that in the notation from (C2),

$$\Gamma_j \subset \bigcup_{i=1}^{v} V(\gamma_i,\bar\delta(\gamma_i)),$$

where $\bar\delta(\gamma) = \min\{\delta(\gamma),\varepsilon(\gamma)\}$. According to the Lebesgue covering lemma there exists a positive number δ_j such that for each $\gamma^*\in\Gamma_j$ one can find an index i for which $V(\gamma^*,\delta_j)$ is a subset of $V(\gamma_i,\bar\delta(\gamma_i))$. Hence we may assume that $\delta_j \le \min\{\bar\delta(\gamma_1),\dots,\bar\delta(\gamma_v)\}$ and therefore in the notation $M_j = \max\{M(\gamma_1),\dots,M(\gamma_v)\}$ the implication

$$\gamma^*\in\Gamma_j,\qquad 0<\delta^*\le\delta_j \quad\Rightarrow\quad \int L(x,V(\gamma^*,\delta^*))\,d\nu(x) \le 1 + M_j\delta^* \tag{2.2}$$

is true. Further, compactness of the set Γ_j implies existence of real numbers $d_j < D_j$ such that

$$\Gamma_j \subset [d_j, D_j]^m. \tag{2.3}$$

For $i_s = 0, \dots, n_j - 1$ let

$$\Gamma_j(n_j, i_1, \dots, i_m) = \Big\{(\gamma_1, \dots, \gamma_m)' \in \Gamma_j; \gamma_s \in \Big[d_j + \frac{(D_j - d_j)}{n_j} i_s,\ d_j + \frac{(D_j - d_j)}{n_j}(i_s + 1)\Big],\ s = 1, \dots, m\Big\}. \tag{2.4}$$

Choosing a $\gamma \in \Gamma_j(n_j, i_1, \dots, i_m)$ if this set is nonempty, we see that $\rho(\gamma, \gamma^*)^2 \le m((D_j - d_j)/n_j)^2$ for all $\gamma^* \in \Gamma_j(n_j, i_1, \dots, i_m)$. Hence making use of (2.2) we obtain that the inequality

$$\int L(x_1, \dots, x_{n_j}, \Gamma_j(n_j, i_1, \dots, i_m))\, d\nu(x_1, \dots, x_{n_j}) \le \Big(1 + M_j \sqrt{m} \frac{(D_j - d_j)}{n_j}\Big)^{n_j} \tag{2.5}$$

holds whenever

$$n_j > \frac{\sqrt{m}(D_j - d_j)}{\delta_j} \tag{2.6}$$

and the set (2.4) is nonempty. The inequalities (2.5), $\log(1 + z) < z$ and the inclusion (2.3) imply that under validity of (2.6),

$$\int L(x(j, n_j), \Gamma_j)\, d\nu(x(j, n_j)) \le n_j^m \Big(1 + M_j \sqrt{m} \frac{(D_j - d_j)}{n_j}\Big)^{n_j} \le n_j^m \exp\big(M_j \sqrt{m}(D_j - d_j)\big). \tag{2.7}$$

From (2.1) and (2.7) we easily get (1.10).

(ii) For every positive number t,

$$P_\theta\Big(\log \frac{L(x_{(n_1,\dots,n_q)}, \Theta)}{L(x_{(n_1,\dots,n_q)}, \theta)} \ge nt\Big) \le \max\Big\{2P_\theta\Big(\log \frac{L(x_{(n_1,\dots,n_q)}, \Theta - \Gamma)}{L(x_{(n_1,\dots,n_q)}, \theta)} \ge 0\Big), 2P_\theta\Big(\log \frac{L(x_{(n_1,\dots,n_q)}, \Gamma)}{L(x_{(n_1,\dots,n_q)}, \theta)} \ge nt\Big)\Big\}. \tag{2.8}$$

Inserting into (2.8) the compact set Γ satisfying (1.9) with $a > A$ and employing (1.10) we obtain that for all $n_1, \dots, n_q$ sufficiently large and for all $t \in [0, A]$, the second term on the right-hand side of (2.8) is larger than the first one, and (1.11) follows from (1.10). □

The constant C in (1.10) can be described by an explicit formula. Making use of (2.7) and (2.1) we see that (1.10) holds with

$$C = \sqrt{m} \sum_{j=1}^{q} M_j (D_j - d_j) \tag{2.9}$$

provided that the inequality (2.6) is fulfilled for $j = 1, \ldots, q$. The quantities appearing in (2.9) are the ones occurring in (2.3) and in the implication (2.2).

3. An application to the exponential distributions. Let

$$K(\theta^*, \theta, p) = \sum_j p_j K(\theta_j^*, \theta_j)$$

and

$$K(\theta_j^*, \theta_j) = E_{\theta_j^*}\left(\log \frac{f(x, \theta_j^*)}{f(x, \theta_j)}\right),$$

with

$$f(x, (\mu, \sigma)) = \frac{1}{\sigma} \exp\left[-\frac{(x - \mu)}{\sigma}\right] \chi_{[\mu, +\infty]}(x)$$

denoting the density of the $E(\mu, \sigma)$ distribution. Further, let

$$\Omega \subset \Theta = \{((\mu_1, \sigma_1), \ldots, (\mu_q, \sigma_q));\ (\mu_j, \sigma_j) \in R \times (0, +\infty),\ \ j = 1, \ldots, q\}$$

and

$$K(\Omega, \theta, p) = \inf\{K(\theta^*, \theta, p);\ \theta^* \in \Omega\},$$
$$K(\theta, \Omega, p) = \inf\{K(\theta, \theta^*, p);\ \theta^* \in \Omega\}.$$

There is a function $g_{n_1, \ldots, n_q}(x_{(n_1, \ldots, n_q)})$ such that for all $\theta \in \Theta$ and $\Omega \subset \Theta$,

$$\log L(x_{(n_1, \ldots, n_q)}, \Omega) = g_{n_1, \ldots, n_q}(x_{(n_1, \ldots, n_q)}) - nK(\hat{\theta}_{n_1, \ldots, n_q}, \Omega, \hat{p}) \tag{3.1}$$

a.e. P_θ, where $\hat{\theta}_{n_1, \ldots, n_q} = (\hat{\theta}_{n_1}, \ldots, \hat{\theta}_{n_q})$, $\hat{\theta}_{n_j} = (\hat{\mu}_j, \overline{x}_j - \hat{\mu}_j)$ is the MLE computed from the jth sample and $\hat{p}_j$ is defined in (1.7). Thus the well-known inequalities

$$\begin{aligned} P_\theta(\hat{\theta} \in \Omega) &\le P_\theta(K(\hat{\theta}, \theta, \hat{p}) \ge K(\Omega, \theta, \hat{p})) \\ &= P_\theta\left(\log \frac{L(x_{(n_1, \ldots, n_q)}, \Theta)}{L(x_{(n_1, \ldots, n_q)}, \theta)} \ge nK(\Omega, \theta, \hat{p})\right) \end{aligned}$$

hold also in this case and we obtain from Theorem 1.1 that under validity of (1.8) given $A > 0$ there is a constant C such that

$$P_\theta(\hat{\theta}_{n_1 \cdots n_q} \in \Omega) \le \exp\left[-nK(\Omega, \theta, \hat{p}) + m \sum_{j=1}^{q} \log n_j + C\right] \tag{3.2}$$

for all $\theta \in \Theta$ and all measurable sets $\Omega \subset \Theta$ for which $K(\Omega, \theta, p) \le A$.

Let Ω_0 be the null hypothesis that $\mu_1 = \cdots = \mu_q$, $\sigma_1 = \cdots = \sigma_q$ and let Ω_1 be the set consisting of the parameters for which $\sigma_1 = \cdots = \sigma_q$. Taking into account the form of the distribution of the differences $x^{(i)} - x^{(i-1)}$ of order statistics for sampling from the $E(\mu, \sigma)$ distribution, denoting $\hat{\mu} = \min\{\hat{\mu}_1, \dots, \hat{\mu}_q\}$ and assuming that $(\mu_i, \sigma_i) = (0, 2)$ for all i, after some computation we get that

$$P\biggl(\sum_{j=1}^{q} n_j(\hat{\mu}_j - \hat{\mu}) < t,\ \hat{\mu} = \hat{\mu}_i\biggr) = \frac{n_i}{n} P\biggl(\sum_{j \neq i} n_j \hat{\mu}_j < t\biggr).$$

Thus $\mathscr{L}(\sum_{j=1}^{q} n_j(\hat{\mu}_j - \hat{\mu})) = \chi^2_{2(q-1)}$ and under the validity of Ω_0, the test statistic

$$T_{n_1,\dots,n_q} = \frac{(n-q)}{(q-1)} \frac{\sum_{j=1}^{q} n_j(\hat{\mu}_j - \hat{\mu})}{\sum_{j=1}^{q} n_j(\bar{x}_j - \hat{\mu}_j)} \tag{3.3}$$

has the F-distribution with $2(q-1), 2(n-q)$ degrees of freedom. We propose to reject Ω_0 in favor of $\Omega_1 - \Omega_0$ if $T_{n_1,\dots,n_q}$ exceeds the quantile $F_{2(q-1),\, 2(n-q),\, 1-\alpha}$ of the F distribution.

To establish the Hodges–Lehmann optimality of this test we assume that (1.8) holds and $\theta \in \Omega_1 - \Omega_0$ and $\alpha \in (0, 1)$ are fixed. Since

$$\log \frac{L(x_{(n_1,\dots,n_q)}, \Omega_1)}{L(x_{(n_1,\dots,n_q)}, \Omega_0)} = n \log\biggl(1 + \frac{q-1}{n-q} T_{n_1,\dots,n_q}\biggr),$$

denoting $w = F_{2(q-1),\, 2(n-q),\, 1-\alpha}$ we obtain from (3.1) and (3.2) that

$$\begin{aligned} \frac{1}{n} \log P_\theta(T_{n_1,\dots,n_q} \le w) &= \frac{1}{n} \log P_\theta(\hat{\theta}_{n_1,\dots,n_q} \in D_{n_1,\dots,n_q}) \\ &\le -K(D_{n_1,\dots,n_q}, \theta, \hat{p}) + o(1), \end{aligned} \tag{3.4}$$

where $D_{n_1,\dots,n_q} = \{\theta^*;\ K(\theta^*, \Omega_0, \hat{p}) - K(\theta^*, \Omega_1, \hat{p}) \le n^{-1} t_{n,\alpha}\}$. Let $\theta^*_{n_1,\dots,n_q} \in D_{n_1,\dots,n_q}$ be such that $K(\theta^*_{n_1,\dots,n_q}, \theta, \hat{p}) - K(D_{n_1,\dots,n_q}, \theta, \hat{p})$ tends to zero. Then $\limsup K(\theta^*_{n_1,\dots,n_q}, \theta, \hat{p}) < +\infty$ and we may assume, that $\theta^*_{n_1,\dots,n_q} \to \theta^* \in \Theta$. Hence $K(\theta^*, \Omega_0, p) - K(\theta^*, \Omega_1, p) = 0$ and θ^* belongs to the set Ω_1^* consisting of the parameters with $\mu_1^* = \cdots = \mu_q^*$. Since for each $\tilde{\theta} \in \Omega_1$ the equality $K(\Omega_1^*, \tilde{\theta}, p) = K(\Omega_0, \tilde{\theta}, p)$ holds, (3.4) implies that

$$\limsup_{n \to \infty} \frac{1}{n} \log P_\theta(T_{n_1,\dots,n_q} \le w) \le -K(\theta^*, \theta, p) \le -K(\Omega_0, \theta, p).$$

Thus the statistics (3.3) are Hodges–Lehmann optimal for testing Ω_0 against $\Omega_1 - \Omega_0$.

REFERENCES

BAHADUR, R. R. (1965). An optimal property of the likelihood ratio statistic. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 13–26. Univ. California Press, Berkeley.

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

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

KOUROUKLIS, S. (1987). Hodges–Lehmann optimality of the likelihood ratio test in regular exponential families of distributions. Technical Report 73, Dept. Statistics, Pennsylvania State Univ.

RUBLÍK, F. (1989). On optimality of the LR tests in the sense of exact slopes. Part I. General case. *Kybernetika* **25** 13–25.

RUBLÍK, F. (1995). On d-optimality of the LR tests. *Kybernetika* **31** 189–206.

INSTITUTE OF MEASUREMENT OF SAS
DÚBRAVSKÁ CESTA 9
84219 BRATISLAVA
SLOVAKIA
E-MAIL: umerrubl@savba.savba.sk