J. Korea Soc. Math. Ed. Series B: The Pure and Applied Mathematics
4, No. 1 (1997) pp. 97–104

CONDITIONAL LARGE DEVIATIONS FOR 1-LATTICE DISTRIBUTIONS

GIE-WHAN KIM

ABSTRACT. The large deviations theorem of Cramér is extended to conditional probabilities in the following sense. Consider a random sample of pairs of random vectors and the sample means of each of the pairs. The probability that the first falls outside a certain convex set given that the second is fixed is shown to decrease with the sample size at an exponential rate which depends on the Kullback-Leibler distance between two distributions in an associated exponential familiy of distributions. Examples are given which include a method of computing the Bahadur exact slope for tests of certain composite hypotheses in exponential families.

1. Introduction.

Suppose that $(X,\ Y)$ has a d-dimensional joint distribution P, where the random vectors X and Y, are p and q dimensional, respectively $(d = p + q)$. P will be assumed to have a finite moment generating function in some open neighborhood of the origin. Consider the exponential family $\{P_\theta\}$ generated by P as follows:

$$dP_\theta(x,y) = e^{\theta_1\cdot x+\theta_2\cdot y-\psi(\theta_1,\theta_2)}dP(x,y)$$

where $\theta = (\theta_1, \theta_2)$ and $\psi(\theta) = \log \int e^{\theta_1\cdot x+\theta_2\cdot y}dP(x,y)$.
Let

$$\Omega = \{\theta \in R^d : \psi(\theta) < \infty\}.$$

We assume that Ω is open so that our exponential family is regular.

Jing and Robinson(1994) studied saddlepoint approximations for the conditional large deviation $P(\bar{X} \geq c|\bar{Y} = y)$ when $p = 1$. In practice we often have interests when $1 < p < k$. For example, a frequently occuring problem in testing hypotheses

Received by the editors May 3, 1997.
1991 *Mathematics Subject Classification.* 60F10.
Key words and phrases. large deviations, exponential families, testing hypotheses, Kullback-Leibler information.

in a multivariate exponential family is to test if $\theta_1 = \cdots = \theta_p = 0$ for some $1 < p < k$ where $\theta_{p+1}, \cdots, \theta_k$ are nuisance parameters. Bahadur efficiency is a popular measuring method of asymptotic efficiencies of competing test statistics. When we have two competing test statistics such as the Wald test and the likelihood ratio test for the above testing problem, Bahadur efficiency is the ratio of two exact slopes of these statistics. To be able to compute these exact slopes, the conditional large deviation of $1 < p < k$ is often needed. In this paper we investigate the conditional large deviation of $P(\bar{X} \notin C|\bar{Y} = y)$ when the distribution P has a 1-lattice distribution. Here, C is a convex subset of R^p, $1 \leq p \leq k$. In section 3, some applications are presented.

2. The Main Results

Suppose that P is a 1-lattice distribution on R^d which we may assume concentrates on those points of R^d having integer coordinates. Recall that the information function for the exponential family can be written as

$$\begin{aligned} I(\theta,\omega) &= E_\theta log\frac{dP_\theta}{dP_\omega}(X,Y) \\ &= (\theta-\omega)\cdot\lambda(\theta) - \psi(\theta) + \psi(\omega) \end{aligned}$$

where

$$\lambda(\theta) = E_\theta(X,Y)$$

Define $\omega(y)$ by

$$I(\omega(y),0) = inf\{I(\theta,0) : \lambda_2(\theta) = y\}$$

For C a convex set in R^p, let $\theta(y)$ be such that

$$I(\theta(y),0) = inf\{I(\theta,0) : \lambda_1(\theta) \notin C,\ \lambda_2(\theta) = y\}$$

Lemma 1. *Let C be a convex Borel set in R^d containing the point $\lambda(\omega)$, and*

$$I(\theta_0,\omega) = inf\{I(\theta,\omega) : \lambda(\theta) \notin C\}$$

Then $\lambda(\theta_0)$ is a boundary point of C, and $\theta_0 - \omega$ is orthogonal to a supporting hyperplane to C at $\lambda(\theta_0)$.

Proof. We first argue that $\lambda(\theta_0)$ is a boundary point of C. Define $I^*(\lambda, \lambda(\omega)) = (\theta - \omega) \cdot \lambda - \psi(\theta) + \psi(\omega)$ where θ is defined by $\lambda(\theta) = \lambda(i.e. I^*(\lambda, \lambda(\omega)) = I(\theta, \omega))$. Noting that

$$\frac{\partial^2 I^*}{\partial \lambda_i \partial \lambda_j} = \sum_{i,j}(\theta)$$

where $\sum(\theta)$ is the covariance matrix corresponding to P_θ, we see that I^* is a strictly convex function of λ as long as θ is interior to Ω. Since $I^*(\lambda, \lambda(\omega))$ has its minimum at $\lambda(\omega)$, it is strictly increasing along rays from $\lambda(\omega)$. Hence, $\lambda(\theta_0)$ must be a boundary point of C.

Now, consider the convex set

$$S = \{\lambda : I^*(\lambda, \lambda(\omega)) \le I^*(\lambda(\theta_0), \lambda(\omega))\}$$

S and C have only boundary points in common. In paticular, $\lambda(\theta_0)$ belongs to both S and C. Thus, there exists a separating hyperplane through $\lambda(\theta_0)$ (which supports both sets), so it has equation

$$(grad I^*)_{\lambda(\theta_0)} \cdot (x - \lambda(\theta_0)) = 0$$

or

$$(\theta_0 - \omega) \cdot (x - \lambda(\theta_0)) = 0$$

because the i^{th} component of $(grad I^*)_{\lambda(\theta_0)}$ is

$$\frac{\partial}{\partial \lambda_i}[(\theta - \omega) \cdot \lambda(\theta) - \psi(\theta) + \psi(\omega)]_{\theta_0} = (\theta_{0i} - \omega_i)$$

which completes the proof.

Lemma 2. $I(\theta(y), 0) - I(\omega(y), 0) = I(\theta(y), \omega(y))$

Proof. By lemma 1, $\omega(y)$ is orthogonal to a hyperplane supporting $\{(s,t) : t = y\}$, and this implies $\omega_1(y) = 0$. Thus, since $\lambda_2(\theta(y)) = \lambda_2(\omega(y)) = y$

$$\begin{aligned} I(\theta(y), 0) - I(\omega(y), 0) &= \theta_1(y) \cdot \lambda_1(\theta(y)) + \theta_2(y) \cdot \lambda_2(\theta(y)) - \psi(\theta(y)) \\ &\quad - \omega_1(y) \cdot \lambda_1(\omega(y)) - \omega_2(y) \cdot \lambda_2(\omega(y)) + \psi(\omega(y)) \\ &= (\theta_1(y) - \omega_1(y)) \cdot \lambda_1(\theta(y)) + (\theta_2(y) - \omega_2(y)) \cdot \lambda_2(\theta(y)) \\ &\quad - \psi(\theta(y)) + \psi(\omega(y)) \\ &= I(\theta(y), \omega(y)). \end{aligned}$$

This finishes the proof of lemma 2.

Theorem 1. *Under the above assumptions, if* $\lambda_1(\omega(y)) \in intC$, *and* $P(\sum_{i=1}^n Y_i = [ny]) > 0$ *for all* n *sufficiently large*

$$\lim_{n\to\infty} \frac{1}{n} log P(\bar{X}_n \notin C | \bar{Y}_n = \frac{[ny]}{n}) = -I(\theta(y), \omega(y)).$$

Proof. Write the conditional probability as

$$P\left(\bar{X}_n \notin C | \bar{Y}_n = \frac{[ny]}{n}\right) = \frac{P(\bar{X}_n \notin C,\ \bar{Y}_n = \frac{[ny]}{n})}{P(\bar{Y}_n = \frac{[ny]}{n})}$$

and treat the numerator and denominator separately.

$$P\left(\bar{Y}_n = \frac{[ny]}{n}\right) = e^{-\omega_2(y)\cdot([ny]-ny)-nI(\omega(y),o)} P_{\omega(y)}\left(V_n = \frac{[ny]-ny}{\sqrt{n}}\right)$$

where

$$V_n = \frac{\sum_{i=1}^n (Y_i - \lambda_2(\omega(y)))}{\sqrt{n}}, \text{ and } \lambda_2(\omega(y)) = y$$

A local limit theorem, such as Corollary 22.3, p.337 of Bhattacharya and Rao (1976), for lattice random vectors yields

$$P_{\omega(y)}(V_n = \frac{[ny]-ny}{\sqrt{n}}) = \frac{\phi^*(0)}{n^{q/2}}(1+o(1))$$

where ϕ^* is a normal density on R^q with covariance matrix $\sum_{22}(\omega(y))$.

For the numerator, first suppose that the convex set C is a half space $a \cdot \bar{x} < c$. Recall that $\theta(y)$ must be a multiple of $(a, 0)$ and $a \cdot \lambda_1(\theta(y)) = c$. Letting $U_n = \sqrt{n}\theta_1(y) \cdot (\bar{X}_n - \lambda_1(\theta(y)))$ we can write

$$P\left(a \cdot \bar{X}_n \geq c,\ \sum_{i=1}^n Y_i = [ny]\right)$$
$$= e^{-nI(\theta(y),0)} \sum_{u\geq 0} e^{-\sqrt{n}u} P_{\theta(y)}\left(U_n = u,\ \sum_{i=1}^n Y_i = [ny]\right)$$

and, using integration by parts

$$P(a \cdot \bar{X}_n \geq c,\ \sum_{i=1}^n Y_i = [ny]) = e^{-nI(\theta(y),0)} \int_0^\infty \sqrt{n} e^{-\sqrt{n}s} P_n^y(0, s) ds.$$

The set function $P_n^y(A)$ is defined by

$$P_n^y(A) = P_{\theta(y)}\left(U_n \in A,\ \sum_{i=1}^{n} Y_i = [ny]\right)$$

Define another set function

$$\Phi^0(A) = \int_A \phi^{**}(x,0)dx$$

where ϕ^{**} is the normal density on $R \times R^q$ with mean 0 and covariance matrix

$$cov_{\theta(y)}(U_1, Y_1)$$

The reader may note that (U, Y) may not have a lattice distribution so that an asymptotic expansion such as used in the density case is not feasible. However, it is possible to make use of a general local limit theorem of Charles Stone (1967) and make an argument paralleling that in Efron and Truax (1968) to establish

$$\lim_{n\to\infty} \frac{\int_0^\infty \sqrt{n}e^{-\sqrt{n}s} P_n^y(0,s)ds}{\frac{1}{n^{q/2}} \int_0^\infty \sqrt{n}e^{-\sqrt{n}s}\Phi^0(0,s)ds} = 1$$

Using the fact that

$$\int_0^\infty \sqrt{n}e^{-\sqrt{n}s}\Phi^0(0,s)ds = \frac{\phi^{**}(0,0)}{\sqrt{n}}(1+o(1))$$

we get

$$P\left(a\cdot \bar{X}_n \geq c | \sum_{i=1}^{n} Y_i = [ny]\right) = \frac{\frac{1}{n^{q/2}}e^{-nI(\theta(y),0)}\frac{\phi^{**}(0,0)}{\sqrt{n}}(1+o(1))}{\frac{1}{n^{q/2}}e^{-nI(\omega(y),0)}\phi^*(0)(1+o(1))}.$$

Then by lemma 2, we have

$$P(a\cdot \bar{X}_n \geq c|\bar{Y}_n = \frac{[ny]}{n}) = e^{-nI(\theta(y),\omega(y))}\frac{\phi^{**}(0,0)}{\sqrt{n}\phi^*(0)}(1+o(1))$$

This finsihes the proof of theorem 1.

3. Applications

1. A frequently occuring problem in testing hypotheses about the parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ in a multivariate exponential family is to test that $\theta_1 = \theta_2 = \cdots = \theta_p = 0$ for some $1 \leq p < k$, where $\theta_{p+1}, \cdots, \theta_k$ are regarded as nuisance parameters. It is well known(Matthes and Truax (1967)) that a complete class of tests for this problem is the class of tests that have acceptance regions of the form: $(\bar{X}, \bar{Y}) \in C$ where $C_y = \{x : (x, y) \in C\}$ is convex for each y. Here, $(\bar{X}, \bar{Y})$ is the mean of a random sample from the exponential family, and $\bar{X}, \bar{Y}$ are the first p and last $k - p$ coordinates of the mean.

If such a test has an exact slope in the sense of Bahadur, then it can be found as

$$\lim_{n\to\infty} \frac{1}{n} log \sup_{\theta\in\Omega_0} P_\theta((\bar{X}, \bar{Y}) \notin C)$$

where $\Omega_0 = \{\theta : \theta_1 = \cdots = \theta_p = 0\}$. For a fixed $\theta_0 \in \Omega_0$

$$P_{\theta_0}((\bar{X}, \bar{Y}) \notin C) = \int P_0(\bar{X} \notin C_y)|\bar{Y} = y) dP_{\theta_0}^{\bar{Y}}(y)$$

By making use of the results in the preceding sections one could approximate the integrand. In fact, under suitable conditions one can show

$$\lim_{n\to\infty} \frac{1}{n} log \sup_{\theta_0\in\Omega_0} \int P_0(\bar{X} \notin C_y|\bar{Y} = y) dP_0^{\bar{Y}}(y) = -\inf_s I(\theta(s), \omega(s))$$

where I is the Kullback-Leibler information for the exponential family and

$$I(\omega(s), 0) = inf\{I(\omega, 0) : \lambda_2(\omega) = s\}$$
$$I(\theta(s), 0) = inf\{I(\theta, 0) : \lambda_1(\theta) \in \partial C_s, \lambda_2(\theta) = s\}$$

We will not pursue this application here. For a different treatment yielding the same result, see Kim (1997).

2. Our final example gives a large deviations result for the hypergeometric distribution. Presumably, one could obtain such a result using Stirling's formula. Suppose X and Z are independent, where X is $b(1, p)$ and Z is $b(k - 1, p)$. Let $(X_i, Z_i), i = 1, 2, \cdots, n$ be independent replications of (X, Z). The conditional distribution of $\sum_{i=1}^n X_i$ given $\sum_{i=1}^n Y_i = y$ (where $Y_i = X_i + Z_i$) is hypergeometric.

$$P(\sum_{i=1}^n X_i = x | \sum_{i=1}^n Y_i = y) = \frac{\binom{n}{x}\binom{m}{y-x}}{\binom{n+m}{y}}$$

where $m = n(k-1)$. By our theorem

$$\frac{1}{n} log P(\sum_{i-1}^{n} X_i \geq [cn] | \sum_{i=1}^{n} Y_i = [yn]) \to -I(\theta(y), \omega(y))$$

if $\lambda_1(\omega(y)) < c < y$. The joint distribution of (X, Z) is a member of the exponential family

$$p_1^x q_1^{1-x} \binom{k-1}{z} p_2^z q_2^{k-1-z}$$

and the corresponding (X, Y) has distribution

$$p_1^x q_1^{1-x} \binom{k-1}{y-x} p_2^{y-x} q_2^{k-1-(y-x)} = \binom{k-1}{y-x} e^{x log \frac{p_1 q_2}{p_2 q_1} + y log \frac{p_2}{q_2}}$$

for $x \leq y \leq x + (k-1)$, $x = 0, 1$ with natural parameters $\theta_1 = log \frac{p_1 q_2}{p_2 q_1}$, $\theta_2 = log \frac{p_2}{q_2}$. It is more convenient to express the information function in terms of the binomial parameters. Without loss we can assume that $p = \frac{1}{2}$, or equivalently, $\theta_1 = \theta_2 = 0$.

$$\begin{aligned} I(\theta, 0) &= E_{p_1, p_2} \left\{ log \frac{p_1^X q_1^{1-X} p_2^{Y-X} q_2^{k-1-Y+X}}{2^{-k}} \right\} \\ &= p_1 log p_1 + q_1 log q_1 + (k-1) p_2 log p_2 + (k-1) q_2 log q_2 + k log 2 \end{aligned}$$

In order to find $\omega(y)$ we set $E_\omega(Y) = y$ or $p_1 + (k-1)p_2 = y$, and minimize the information function. It is a straightforward calculation to fine

$$\lambda_1(\omega(y)) = \frac{y}{k}$$

and

$$\begin{aligned} I(\theta(y), \omega(y)) &= I(\theta(y), 0) - I(\omega(y), 0) \\ &= c\ log c + (1-c) log(1-c) + (y-c) log \frac{y-c}{k-1} \\ &\quad + (k-1-y+c) log(1 - \frac{y-c}{k-1}) - y\ log \frac{y}{k} - (k-y) log(1 - \frac{y}{k}) \end{aligned}$$

References

1. Bahadur, R.R., *Stochastic comparison of tests*, Ann. Math. Statist. **31** (1960), 276-295.
2. Bahadur, R.R. and Rao, R.Ranga, *On deviations of the sample mean*, Ann. Math. Statist. **31** (1960), 1015-1027.
3. Bhattacharya, R.N. and Rao, R.Ranga, *Normal Approximation and Asymptotoc Expansions*, Wiley, New York (1976).

4. Jing, B. and Robinson, J, *Saddlepoint Approximations for Marginal and Conditional Probabilities of Transformed Variables*, Ann. of Statist (1994).
5. Cramér, H, *Sur un nouveau théorème-limite de la théorie de la probabilités*, Actualités Scientifiques et Industrielles **736** (1938), Paris.
6. Efron, Bradley and Truax, Donald, *Large deviations theory in exponential families*, Ann. Math. Statist **39** (1968), 1402-1424.
7. Kim, Gie-Whan, *Exact slopes of test statistics for the multivariate exponential family*, Scandinavian Journal of Statist (1997).
8. Matthes, T.K. and Truax, D.R., *Tests of composite hypotheses for the multivariate exponential family*, Ann. Math. Statist **38** (1967), 681-697.
9. Stone, Charles, *On local and ratio limit theorems*, Proc. Fifth Berkeley Symp. Math. Statis. Probab., **II** (1967), pt.2, 217-224. Univ. California Press, Berkeley.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OREGON, EUGENE, OREGON 97402, U.S.A.