ON PROBABILITIES OF LARGE DEVIATIONS

WASSILY HOEFFDING
UNIVERSITY OF NORTH CAROLINA

1. Summary

The paper is concerned with the estimation of the probability that the empirical distribution of n independent, identically distributed random vectors is contained in a given set of distributions. Sections 1–3 are a survey of some of the literature on the subject. In section 4 the special case of multinomial distributions is considered and certain results on the precise order of magnitude of the probabilities in question are obtained.

2. The general problem

Let $X_1, X_2, \cdots$ be a sequence of independent m-dimensional random vectors with common distribution function (d.f.) F. If we want to obtain general results on the behavior of the probability that $X^{(n)} = (X_1, \cdots, X_n)$ is contained in a set A^* when n is large, we must impose some restrictions on the class of sets. One interesting class consists of the sets A^* which are symmetric in the sense that if $X^{(n)}$ is in A^*, then every permutation $(X_{j_1}, \cdots, X_{j_n})$ of the n component vectors of $X^{(n)}$ is in A^*. The restriction to symmetric sets can be motivated by the fact that under our assumption all permutations of $X^{(n)}$ have the same distribution. Let $F_n = F_n(\cdot|X^{(n)})$ denote the empirical d.f. of $X^{(n)}$. The empirical distribution is invariant under permutations of $X^{(n)}$, and for any symmetric set A^* there is at least one set A in the space $\mathcal{G}$ of m-dimensional d.f.'s such that the events $X^{(n)} \in A^*$ and $F_n(\cdot|X^{(n)}) \in A$ are equivalent. The latter event will be denoted by $F_n \in A$ for short. Thus when we restrict ourselves to symmetric sets, we may as well consider the probabilities $P\{F_n \in A\}$, where $A = A_n$ may depend on n. (It is understood that $A \subset \mathcal{G}$ is such that the set $\{x^{(n)}|F_n(\cdot|x^{(n)}) \in A\}$ is measurable.) Since F_n converges to F in a well-known sense (Glivenko-Cantelli theorem), we may say that $P\{F_n \in A_n\}$ is the probability of a large deviation of F_n from F if F is not in A_n and not "close" to A_n, implying that $P\{F_n \in A_n\}$ approaches 0 as $n \to \infty$. For certain classes of sets A_n estimates of $P\{F_n \in A_n\}$

This research was supported in part by the Mathematics Division of the Air Force Office of Scientific Research. Part of the work was done while the author was a visiting professor at the Research and Training School of the Indian Statistical Institute in Calcutta under the United Nations Technical Assistance Program.

(some of which are mentioned below) have been obtained which hold uniformly for "large" and for "small" deviations.

For any two d.f.'s F and G in $\mathcal{G}$ let μ be some sigma-finite measure which dominates the two distributions (for instance, $d\mu = d(F + G)$), and let f and g be the corresponding densities, $dF = f\,d\mu$, $dG = g\,d\mu$. Define

$$I(G, F) = \int (\log (g/f))g\,d\mu, \tag{1}$$

with the usual convention that the integrand is 0 whenever $g = 0$. (The value of $I(G, F)$ does not depend on the choice of μ.) We have $0 \le I(G, F) \le \infty$; $I(G, F) = 0$ if and only if $G = F$; $I(G, F) < \infty$ only if F dominates G. Let

$$I(A, F) = \inf_{G \in A} I(G, F), \tag{2}$$

$I(A, F) = +\infty$ if A is empty. Sanov [12] has shown that under certain restrictions on A_n,

$$P\{F_n \in A_n\} = \exp\{-nI(A_n, F) + o(n)\}. \tag{3}$$

If F is discrete and takes only finitely many values, the distribution of F_n may be expressed in terms of a multinomial distribution. In this case the estimate (3), with $o(n)$ replaced by $O(\log n)$, holds under rather mild restrictions on A_n (see [5] and section 4). In [12] (where only sets A independent of n and one-dimensional distributions F are considered) Sanov obtains (3) for a certain class of sets A such that $P\{F_n \in A\}$ can be approximated by multinomial probabilities.

Some necessary conditions for (3) to be true are easily noticed. Let $\mathcal{G}_n(F)$ denote the set of all $G \in \mathcal{G}$ such that nG is integer-valued and $\int_E dF = 0$ implies $\int_E dG = 0$ for every open set $E \subset R^m$. Then $F_n \in \mathcal{G}_n(F)$ with probability one and $P\{F_n \in A\} = P\{F_n \in A \cap \mathcal{G}_n(F)\}$. Let $\mathcal{G}(F)$ denote the set of all G which are dominated by F. Then $I(A, F) < \infty$ only if $A \cap \mathcal{G}(F)$ is not empty. Hence, $\exp\{-nI(A, F)\}$ can be a nontrivial estimate of $P\{F_n \in A\}$ only if both $A \cap \mathcal{G}_n(F)$ and $A \cap \mathcal{G}(F)$ are nonempty. If F is discrete, then $\mathcal{G}_n(F) \subset \mathcal{G}(F)$; if F takes only finitely many values, (3) is always true for $A^{(n)} = A_n \cap G_n(F)$ (see (48), section 4), and (3) holds if $I(A_n^{(n)}, F) - I(A_n, F)$ is not too large. If F is absolutely continuous with respect to Lebesgue measure, then $\mathcal{G}_n(F)$ and $\mathcal{G}(F)$ are disjoint; for (3) to be true and nontrivial, A_n must, as a minimum requirement, contain both values of F_n (which are discrete) and d.f.'s which are dominated by F (hence also by the Lebesgue measure).

In the following two sections the approximation (3) will be related to known results for certain classes of sets, which give more precise estimates of the probability.

3. Half-spaces

Let φ be a real-valued measurable function on R^m, and let

$$H = H(\varphi) = \left\{G \middle| \int \varphi\,dG \ge 0\right\} \tag{4}$$

be the set of all $G \in \mathcal{G}$ such that $\int \varphi \, dG$ is defined and nonnegative. The set H may be called a half-space in $\mathcal{G}$. The asymptotic behavior of $P\{F_n \in H) = P\{\sum_{j=1}^n \varphi(X_j) \geq 0\}$ has been studied extensively. To relate these results to the estimate (3), we first prove the following lemma. We shall write $G[B]$ for $\int_B dG$ and $G[\varphi \in E]$ for $G[\{x|\varphi(x) \in E\}]$.

LEMMA 1. *Let* $H = \{G|\int \varphi \, dG \geq 0\}$, $M(t) = \int \exp(t\varphi) \, dF$.

(A) *We have*

$$I(H, F) = -\log \inf_{t \geq 0} M(t); \tag{5}$$

(B) $0 < I(H, F) < \infty$ *if* $\int \varphi \, dF < 0$, $F[\varphi \geq 0] > 0$, $M(t) < \infty$ *for some* $t > 0$;

(B_1) *if, in addition,* $M'(t^*-) > 0$, *where* $t^* = \sup\{t|M(t) < \infty\}$ *and* $M'(t) = dM(t)/dt$, *then* $\inf_{t \geq 0} M(t) = M(t_\varphi)$, *where* $t_\varphi > 0$ *is the unique root of* $M'(t) = 0$;

(C) $I(H, F) = 0$ *if* $\int \varphi \, dF \geq 0$ *or* $M(t) = \infty$ *for all* $t > 0$;

(D) $I(H, F) = \infty$ *if* $F[\varphi \geq 0] = 0$.

PROOF. If $G \in H$ and $I(G, F) < \infty$, then for $t \geq 0$,

$$\begin{aligned} -I(G, F) &\leq t \int \varphi \, dG - I(G, F) = \int \log(\exp(t\varphi) f g^{-1}) g \, d\mu \\ &\leq \log \int \exp(t\varphi) f \, d\mu = \log M(t) \end{aligned} \tag{6}$$

by Jensen's inequality. Hence, $I(H, F) \geq -\log \inf_{t \geq 0} M(t)$. The equality sign holds in both inequalities in (6) if $\int \varphi \, dG = 0$ and $\exp(t\varphi) f g^{-1} = \text{const. a.e. } (F)$. If $M(t) < \infty$ and $M'(t)$ exists, these conditions are equivalent to $dG = \exp(t\varphi) \, dF/M(t)$ and $M'(t) = 0$. Under the hypothesis of (B), $M'(t)$ exists for $0 < t < t^* = \sup\{t|M(t) < \infty\}$, $M'(0+) < 0$, and $M'(t)$ is increasing. Hence if $M'(t^*-) > 0$, then the root t_φ of $M'(t) = 0$ is unique and positive, and $M(t_\varphi) < 1$. This implies (5), (B), and (B_1) under the condition of (B_1). In particular, if $t^* = \infty$ and the conditions of (B) hold, then that of (B_1) also holds. Next, under the hypothesis of (B), if $F[\varphi > 0] = 0$, then $0 < F[\varphi = 0] < 1$, $\inf_{t \geq 0} M(t) = M(\infty) = F[\varphi = 0]$, and the distribution G with $G[\varphi = 0] = 1$ is in H and $I(G, F) = -\log F[\varphi = 0]$. The remaining case of part (B) is where $t^* < \infty$, $M'(t) < 0$ for $t < t^*$, and $F[\varphi > 0] > 0$. Then $\inf_{t \geq 0} M(t) = M(t^*) < 1$, and we must show that $I(H, F) = -\log M(t^*)$. Let $M_c(t) = \int_{\varphi < c} \exp(t\varphi) \, dF$, which is finite for $t \geq 0$ and $c > \infty$. For c large enough there is a unique number $t(c) > t^*$ such that $M_c'(t(c)) = 0$. It is easy to show that $t(c) \to t^*$ and $M_c(t(c)) \to M(t^*)$ as $c \to \infty$. Let G_c be the d.f. defined by $dG_c = \exp(t(c)\varphi) \, dF/M_c(t(c))$ for $\varphi < c$, $G_c[\varphi \geq c] = 0$. Then $G_c \in H$ and $I(G_c, F) = -\log M_c(t(c)) \to -\log M(t^*)$ as $c \to \infty$, so that $I(H, F) = -\log M(t^*)$. The statements (5), (C), and (D) in the cases $\int \varphi \, dF \geq 0$ and $F[\varphi \geq 0] = 0$ are easily verified, and the part of (C) where $M(t) = \infty$ for all $t > 0$ and $\int \varphi \, dF < 0$ is handled exactly like the last case of part (B), completing the proof.

We have the elementary and well-known inequality

$$P\{F_n \in H\} = P\left\{\sum_{j=1}^n \varphi(X_j) \geq 0\right\} \leq \inf_{t \geq 0} M(t)^n = \exp\{-nI(H, F)\}. \tag{7}$$

Equality is attained only in the trivial cases $F[\varphi \leq 0] = 1$ and $F[\varphi \geq 0] = 1$.

Let $H_c = H(\varphi - c) = \{G | \int \varphi\, dG \geq c\}$, where c is a real number. Then $P\{F_n \in H_c\} = P\{\sum_{j=1}^n \varphi(X_j) \geq nc\}$. If $M(t) < \infty$ for $0 < t < t^*$, and $\int \varphi\, dF < c < L'(t^*-)$, where $L(t) = \log M(t)$, then, by lemma 1,

$$I(H_c, F) = ct(c) - L(t(c)) = I^*(c), \tag{8}$$

say, where $t(c) = t_{\varphi - c}$ is defined by $L'(t(c)) = c$.

A theorem of Cramér [3] as sharpened by Petrov [8] can be stated as follows. Suppose that

$$\int \varphi\, dF = 0, \qquad F[\varphi \neq 0] > 0, \qquad M(t) < \infty \quad \text{if} \quad |t| < t_0 \tag{9}$$

for some $t_0 > 0$. Then for $c = c_n > \alpha n^{-1/2}$, $(\alpha > 0)$, $c = o(1)$ as $n \to \infty$, we have

$$P\left\{\sum_{j=1}^n \varphi(X_j) \geq nc\right\} = b_n(c) \exp\{-nI^*(c)\}(1 + O(c)), \tag{10}$$

where, with $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x \exp(-y^2/2)\, dy$,

$$b_n(c) = (1 - \Phi(x)) \exp(-x^2/2), \qquad x = n^{1/2}c/\sigma, \qquad \sigma^2 = \int \varphi^2\, dF. \tag{11}$$

(Usually the theorem is stated in terms of an expansion of $I^*(\sigma n^{-1/2}x)$ in powers of $n^{-1/2}x$.) Petrov [8] also shows that for any $\epsilon > 0$, equation (10) with $O(c)$ replaced by $r\sigma\epsilon$ holds uniformly for $0 < c < \sigma\epsilon$, where $|r|$ does not exceed an absolute constant. (Compare also the earlier paper of Feller [4].)

Bahadur and Rao [1] have obtained an asymptotic expression for the probability in (10) when c is fixed. It implies that if conditions (9) are satisfied, then for c fixed, $o < c < L'(t^*-)$

$$P\left\{\sum_{j=1}^n \varphi(X_j) \geq nc\right\} \asymp n^{-1/2} \exp\{-nI^*(c)\}. \tag{12}$$

(The notation $a_n \asymp b_n$ means that a_n and b_n are of the same order of magnitude, that is, a_n/b_n is bounded away from zero and infinity from some n on.)

From (11) it is seen that $b_n(c) \asymp x^{-1} \asymp c^{-1}n^{-1/2}$ if $c > \alpha n^{-1/2} (\alpha > 0)$. Hence, the quoted results imply the following uniform estimate of the order of magnitude of the probability under consideration. Let α and β be positive numbers such that $\beta < L'(t^*-)$. If conditions (9) are satisfied, then

$$P\left\{\sum_{j=1}^n \varphi(X_j) \geq nc\right\} \asymp c^{-1}n^{-1/2} \exp\{-nI^*(c)\} \tag{13}$$

uniformly for $\alpha n^{-1/2} < c < \beta$. This also follows from ([1], inequality (57)).

In the case $c \to \infty$, A. V. Nagaev [15] obtained, under certain restrictions on the (assumed) probability density of $\varphi(X_1)$, an asymptotic expression for the probability in (12), which is identical with the leading term of the expansion derived in [1] for c fixed.

Lemma 1 shows that $\exp\{-nI(H, F)\}$ does not approximate $P\{F_n \in H\}$ if $M(t) = \infty$ for all $t > 0$. In this case, S. V. Nagaev [16] showed, under a smoothness condition on $F_\varphi(x) = P\{\varphi(X_1) < x\}$, that $P\{\sum_{j=1}^n \varphi(X_j) \geq nc\} \sim n(1 - F_\varphi(nc))$ if nc increases rapidly enough (see also Linnik [14]), whereas Petrov [9], extending the results of Linnik [7], obtained asymptotic expressions for this probability, of the form (10) but with $I^*(c)$ replaced by a partial sum of its expansion in powers of c, under the assumption that nc does not grow too fast.

For certain sets A the results on half-spaces enable us to obtain upper and/or lower bounds for $P\{F_n \in A\}$ of the form (3). If A is any subset of $\mathcal{G}$, it follows from the definition of $I(A, F)$ that $A \subset B = \{G | I(G, F) \geq I(A, F)\}$. Suppose that $0 < I(A, F) < \infty$ and that there is a $G_0 \in A$ such that $I(G_0, F) = I(A, F)$. If $I(G, F)$ and $I(G, G_0)$ are finite, we have $I(G, F) = \int \log (g_0/f)\, dG + I(G, G_0)$, where $f = dF/d\mu$, $g_0 = dG/d\mu$. Hence, the half-space $H = \{G | \int \log (g_0/f)\, dG \geq I(A, F)\}$ is a subset of B, and $I(H, F) = I(A, F)$. In general, H neither contains nor is contained in A.

Suppose that A contains a half-space H such that $I(H, F) = I(A, F)$. For example, if A is the union of a family of half-spaces $H(\varphi)$, $\varphi \in \Phi$ (so that the complement of A is convex), it is easily seen that $I(A, F) = \inf \{I(H(\varphi), F), \varphi \in \Phi\}$. If the infimum is attained in Φ, the stated assumption is satisfied. Then we have the lower bound $P\{F_n \in A\} \geq P\{F_n \in H\}$, which, under appropriate conditions, can be estimated explicitly, as in (10) or (12), where $I^*(c) = I(A, F)$. If A is contained in a half-space H and $I(H, F) = I(A, F)$, we have analogous upper bounds, including $P\{F_n \in A\} \leq \exp\{-nI(A, F)\}$.

Now suppose that the set A is contained in the union of a finite number $k = k(n)$ of half-spaces H_i, $i = 1, \cdots, k$. Then (using (7))

$$(14) \qquad \begin{aligned} P\{F_n \in A\} &\leq \sum_{i=1}^k P\{F_n \in H_i\} \leq \sum_{i=1}^k \exp\{-nI(H_i, F)\} \\ &\leq k \exp\{-n \min_i I(H_i, F)\}. \end{aligned}$$

If $\min I(H_i, F)$ is close to $I(A, F)$ and $k = k(n)$ is not too large, even the crudest of the three bounds in (14) may be considerably better than the upper bound implied by (3). The following example serves as an illustration.

Let

$$(15) \qquad A = \{G | \sup_{x \in R^m} |G(x) - F(x)| \geq c\}, \qquad 0 < c < 1.$$

The set A is the union of the half-spaces $H_x^+ = \{G | G(x) - F(x) \geq c\}$, $H_x^- = \{G | F(x) - G(x) \geq c\}$, $x \in R^m$. Sethuraman [13] has shown that the estimate (3) holds in the present case with c fixed, and for more general unions of half-spaces.

It follows from lemma 1 by a simple calculation that $I(H_x^+, F) = J(F(x), c)$ and $I(H_x^-, F) = J(1 - F(x), c)$, where

$$(16) \qquad \begin{aligned} J(p, c) = {} & (p + c) \log ((p + c)/p) \\ & + (1 - p - c) \log ((1 - p - c)/(1 - p)) \end{aligned}$$

if $0 < p < 1 - c$, $J(1 - c, c) = -\log (1 - c)$, $J(p, c) = \infty$ if $p = 0$ or $p > 1 - c$.

I shall assume for simplicity that the one-dimensional marginal d.f.'s of F are continuous. Then $F(x)$ takes all values in $(0, 1)$, and we have

$$I(A, F) = \min_p J(p, c) = J(p(c), c) = K(c), \tag{17}$$

say, where $p(c)$ is the unique root in $(0, 1 - c)$ of $\partial J(p, c)/\partial p = 0$. It is easy to show that $((1 - c)/2) < p(c) < \min(\frac{1}{2}, 1 - c)$. For $K'(c) = dK(c)/dc$ we find

$$K'(c) = cp^{-1}(c)[1 - p(c)]^{-1} < 4c/(1 - c^2). \tag{18}$$

For any x with $F(x) = p(c)$ we have

$$P\{F_n \in A\} \geq P\{F_n \in H_x^+\} \geq \binom{n}{r} p(c)^r(1 - p(c))^{n-r}, \tag{19}$$

where $r = r(n, c)$ is the integer defined by $r \geq n(p(c) + c) > r - 1$. An application of Stirling's formula shows that this lower bound is greater than $C_1 n^{-1/2} \exp\{-nJ(p(c), c_n)\}$, where C_1 is a positive constant independent of c and $c_n = (r/n) - p(c) = c + \theta/n$, $0 \leq \theta < 1$. Hence it can be shown that for every $\epsilon > 0$ there is a positive constant C_2 which depends only on ϵ such that for $0 < c < 1 - \epsilon$,

$$P\{\sup_{x \in R^m} |F_n(x) - F(x)| \geq c\} \geq C_2 n^{-1/2} \exp\{-nK(c)\}. \tag{20}$$

Now let k be a positive integer. Since the marginal d.f.'s $F^{(i)}(x^{(i)})$ of $F(x) = F(x^{(1)}, \cdots, x^{(m)})$ are continuous, there are numbers $a_j^{(i)}$,

$$-\infty = a_0^{(i)} < a_1^{(i)} < \cdots < a_{k-1}^{(i)} < a_k^{(i)} = +\infty, \tag{21}$$

$i = 1, \cdots, m$, such that $F^{(i)}(a_j^{(i)}) = j/k$ for all i, j. If

$$a_{j_i-1}^{(i)} \leq x^{(i)} < a_{j_i}^{(i)} \qquad \text{for} \quad i = 1, \cdots, m \tag{22}$$

then

$$G(x) - F(x) \leq G(a) - F(a) + m/k, \tag{23}$$

where $a = (a_{j_1}^{(1)}, \cdots, a_{j_m}^{(m)})$, $1 \leq j_i \leq k$, and we have a similar upper bound for $F(x) - G(x)$. Hence the set A is contained in the union of the $2k^m$ half-spaces $\{G|G(a) - F(a) \geq c - m/k\}$, $\{G|F(a) - G(a) \geq c - m/k\}$, corresponding to the k^m values a. If $c - m/k > 0$, we have for each of these half-spaces H the inequality $I(H, F) \geq K(c - m/k)$. Hence, by (14),

$$P\{F_n \in A\} \leq 2k^m \exp\{-nK(c - m/k)\}. \tag{24}$$

We have $K(c - m/k) = K(c) - (m/k)K'(c - \theta m/k)$, $0 < \theta < 1$. With (18) this implies

$$K(c - m/k) > K(c) - (m/k)4c/(1 - c^2), \tag{25}$$

$$P\{F_n \in A\} < 2(k \exp\{4nc(1 - c^2)^{-1}k^{-1}\})^m \exp\{-nK(c)\}. \tag{26}$$

If we choose k so that $k - 1 \leq 4nc(1 - c^2)^{-1} \leq k$ and take account of the assumption $c > mk^{-1}$, we obtain

$$P\{\sup_{x \in R^m} |F_n(x) - F(x)| \geq c\} < 2e^m\{4cn/(1 - c^2) + 1\}^m \exp\{-nK(c)\} \tag{27}$$

if $4c^2 n > m(1 - c^2)$.

For c fixed the bound is of order $n^m \exp\{-nK(c)\}$. The power n^m can be reduced by using the closer bounds in (14). An upper bound of a different form for the probability in (27) has been obtained by Kiefer and Wolfowitz [6].

4. Sums of independent random vectors

Let $\varphi = (\varphi_1, \cdots, \varphi_k)$ be a measurable function from R^m to R^k and consider the set

$$A = \left\{G \mid \int \varphi \, dG \in D\right\}, \tag{28}$$

where D is a k-dimensional Borel set. Then $P\{F_n \in A\}$ is the probability that the sum $n^{-1} \sum_{j=1}^n \varphi(X_j)$ of n independent, identically distributed random vectors is contained in the set D. We have

$$I(A, F) = \inf_{s \in D} I(A(s), F), \qquad A(s) = \left\{G \mid \int \varphi \, dG = s\right\}. \tag{29}$$

For $t \in R^k$ let $M(t) = \int \exp{(t, \varphi)} \, dF$, $L(t) = \log M(t)$, where $(t, \varphi) = \sum_{i=1}^k t_i \varphi_i$. Let Θ denote the set of points $t \in R^k$ for which $M(t) < \infty$. Suppose that the set Θ_0 of inner points of Θ is not empty. The derivatives $L_i'(t) = \partial L(t)/\partial t_i$ exist in Θ_0. Let Ω_0 denote the set of points $L'(t) = (L_1'(t), \cdots, L_k'(t))$, $t \in \Theta_0$. The following lemma, in conjunction with (29), is a partial extension of lemma 1.

LEMMA 2. *If $s \in \Omega_0$, then*

$$I(A(s), F) = (t(s), s) - L(t(s)) = - \min_{t \in R^k} [L(t) - (t, s)], \tag{30}$$

where $t(s)$ satisfies the equation $L'(t(s)) = s$. Also, $I(A(s), F) = I(G_s, F)$, where G_s is the d.f. in $A(s)$ defined by

$$dG_s = \exp{\{(t(s), \varphi)\}} \, dF/M(t(s)). \tag{31}$$

PROOF. If $G \in A(s)$, we find as in (6) that $-I(G, F) \le L(t) - (t, s)$ for all $t \in R^k$, with equality holding only if

$$dG = \exp{\{(t, \varphi)\}} \, dF/M(t). \tag{32}$$

The d.f. G defined by (32) is in $A(s)$ if and only if $\int \varphi \{\exp{(t, \varphi)}\} \, dF/M(t) = s$ which for $t \in \Theta_0$ is equivalent to $L'(t) = s$. Since $s \in \Omega_0$, there is at least one point $t(s) \in \Theta_0$ which satisfies this equation. The lemma follows. (If the distribution of the random vector $\varphi(X_1)$ is concentrated on a hyperplane in R^k, the solution $t(s)$ of $L'(t) = s$ is not unique; but the distribution G_s can be shown to be the only $G \in A(s)$ for which $I(G, F) = I(A(s), F)$.)

It is seen from (30) and (31) that if $s \in \Omega_0$, then

$$dF(x) = \exp{\{-I(A(s), F) - (t(s), \varphi(x) - s)\}} \, dG_s(x), \tag{33}$$

$$\prod_{j=1}^n dF(x_j) = \exp{\left\{-nI(A(s), F) - n\left(t(s), \int \varphi \, dF_n - s\right)\right\}} \prod_{j=1}^n dG_s(x_j). \tag{34}$$

(Here the same notation F_n is used for the value $F_n(\cdot | x^{(n)})$ as for the random function $F_n(\cdot | X^{(n)})$.) Hence the distribution of the sum $\int \varphi \, dF_n$ can be symbolically expressed in the form

(35) $$\int_{\{\int \varphi\, dF_n = s\}} \prod_{j=1}^{n} dF(x_j) = \exp\{-nI(A(s), F)\} \int_{\{\int \varphi\, dF_n = s\}} \prod_{j=1}^{n} dG_s(x_j)$$

for values $s \in \Omega_0$. Here $\{\int \varphi\, dF_n = s\}$ is a shortcut notation for $\{|\int \varphi\, dF_n - s| < \epsilon\}$, $\epsilon \to 0$, and a term which is negligible for $\epsilon \to o$ is suppressed. The integral on the right is the value at s of the distribution of $\int \varphi\, dF_n$ when the $X_1, \cdots, X_n$ have the common distribution G_s, in which case $s = \int \varphi\, dG_s$ is the expected value of $\int \varphi\, dF_n$. The higher moments of this distribution are finite, and the known results on the approximation of the density of a sum of independent random vectors in the center of the distribution can be used to approximate the density (on the left in (35)) at points remote from the center. This, in turn, can be used to approximate $P\{\int \varphi\, dF_n \in D\}$ at least for $D \subset \Omega_0$. This approach has been used by Borovkov and Rogozin [2] to derive an asymptotic expansion of the probability $P\{\int \varphi\, dF_n \in D_n\}$ for an extensive class of sets D_n under the assumption that the distribution of $\int \varphi\, dF_n$ is absolutely continuous with respect to Lebesgue measure in R^k for some n.

Borovkov and Rogozin make the following assumptions concerning D_n. Let $-\psi_n$ denote the essential infimum relative to k-dimensional Lebesgue measure of $I(A(s), F)$ for $s \in D_n$. (Thus $\psi_n = -I(A_n, F)$ where $A_n = \{G | \int \varphi\, dG \in D_n^*\}$ and D_n^* differs from D_n by a set of Lebesgue measure 0.) Let Θ_f be a compact subset of Θ_0 and $\Phi = \{L'(t) | t \in \Theta_f\}$.

ASSUMPTION (A). *For some* $\delta > 0$, $D_n \cap \{s | I(A(s), F) < -\psi_n + \delta\} \in \Phi$.

ASSUMPTION (B). *There is a union U of finitely many half-spaces in R_k such that*

(36) $$D_n \cap \{s | I(A(s), F) > -\psi_n + \delta\} \subset U \subset \left\{s | I(A(s), F) > -\psi_n + \frac{\delta}{2}\right\}.$$

Under these assumptions the leading term of the asymptotic expansion obtained in [2] is

(37) $$P\left\{\int \varphi\, dF_n \in D_n\right\} \sim (2\pi)^{-k/2} n^{k/2} \exp(n\psi_n) \int_0^{\delta} e^{-nu} \varphi_n(u)\, du,$$

where

(38) $$\varphi_n(u) = \int_{D_n \cap \Gamma(-\psi_n - u)} |\textstyle\sum (s)|^{-1/2}\, ds, \qquad \Gamma(-c) = \{s | I(A(s), F) = c\},$$

$|\sum (s)|$ is the determinant of the covariance matrix of $\varphi(X_1)$ when X_1 has the distribution G_s, and the last integral is extended over the indicated surface.

It should be feasible to obtain an analogous expansion for the case of lattice-valued random vectors. An extension of the Euler-Maclaurin sum formula to the case of a function of several variables due to R. Ranga Rao (in a Ph.D. dissertation which is unpublished at this writing; compare [10]), would be useful here. The order of magnitude of the probability $P\{\int \varphi\, dF_n \in D_n\}$ for a fairly extensive class of sets D_n can be determined in a rather simple way, as is shown in section 4 for the multinomial case. Richter [11] derived an estimate of $P\{\int \varphi\, dF_n \in D\}$ for a special class of sets D in the lattice vector case as well as in the absolutely continuous case; it is akin to the Cramér-Petrov estimate for the one-dimensional case but seems to have no simple relation to the Sanov-type estimate (3).

The preceding discussion has an interesting statistical interpretation. Lemma 2 shows that if $s \in \Omega_0$, then the infimum $I(A(s), F)$ is attained in the "exponential" subclass of $\mathcal{G}$ which consists of the distributions G defined by (32). Suppose that $F = F_\theta$ is a member of the class $\{F_\theta, \theta \in \Theta\}$, $dF_\theta = f_\theta \, d\nu$, where

$$f_\theta(x) = \exp \{(\theta, \varphi(x)) - L(\theta)\}, \tag{39}$$

ν is a sigma-finite measure on the m-dimensional Borel sets, φ a function from R^m to R^k, and Θ is the set of points $\theta \in R^k$ for which $\exp L(\theta) = \int \exp (\theta, \varphi) \, d\nu$ is finite. (If the null vector 0 is in Θ, which could be assumed with no loss of generality, then $d\nu = dF_0$.) Let $f_{\theta,n} = f_{\theta,n}(x^{(n)})$ be the density of $X^{(n)}$, so that

$$f_{\theta,n} = \exp n\left\{\left(\theta, \int \varphi \, dF_n\right) - L(\theta)\right\}. \tag{40}$$

Here $\int \varphi \, dF_n$ is a sufficient statistic and it is natural to restrict attention to sets A of the form (28). We have $\int \varphi \, dF_\theta = L'(\theta)$ for $\theta \in \Theta_0$, and

$$I(F_{\theta'}, F_\theta) = (\theta' - \theta, L'(\theta')) - L(\theta') + L(\theta) = I^*(\theta', \theta), \tag{41}$$

say, for $\theta' \in \Theta_0$, $\theta \in \Theta$. From (30) with $F = F_\theta$ we have $I(A(s), F_\theta) = I^*(\theta', \theta)$, where $s = L'(\theta')$.

A maximum likelihood estimator of θ is a function $\hat{\theta}_n$ from R^{mn} into Θ such that $\hat{\theta}_n(x^{(n)})$ maximizes $f_{\theta,n}(x^{(n)})$. If $\int \varphi \, dF_n \in \Omega_0$, then $\hat{\theta}_n$ is a root of $L'(\theta) = \int \varphi \, dF_n$, and we have

$$\max_\theta f_{\theta,n} = f_{\hat{\theta}_n,n} = \exp n\{(\hat{\theta}_n, L'(\hat{\theta}_n)) - L(\hat{\theta}_n)\}. \tag{42}$$

Hence,

$$f_{\theta,n} = f_{\hat{\theta}_n,n} \exp \{-nI^*(\hat{\theta}_n, \theta)\}. \tag{43}$$

Equation (43), which is related to (35), shows that $f_{\theta,n}$ depends on θ only through $I^*(\hat{\theta}_n, \theta)$. Note that the likelihood ratio test for testing the simple hypothesis $\theta = \theta'$ against the alternatives $\theta \neq \theta'$ rejects the hypothesis if $I^*(\hat{\theta}_n, \theta')$ exceeds a constant. For the special case where the distribution of $\int \varphi \, dF_n$ is multinomial the author has shown in [5] that the likelihood ratio test has certain asymptotically optimal properties.

5. Multinomial probabilities

The case where X_1 takes only finitely many values can be reduced to the case where X_1 is a vector of k components and takes the k values $(1, 0, \cdots, 0)$, $(0, 1, 0, \cdots, 0)$, $\cdots$, $(0, \cdots, 0, 1)$ with respective probabilities $p_1, p_2, \cdots, p_k$ whose sum is 1. The sum $nZ^{(n)} = X_1 + \cdots + X_n$ takes the values $nz^{(n)} = (n_1, \cdots, n_k)$, $n_i \geq 0$, $\sum_{i=1}^k n_i = n$, and we have

$$P\{Z^{(n)} = z^{(n)}\} = n! \left(\prod_{i=1}^k n_i!\right)^{-1} \prod_{i=1}^k p_i^{n_i} = P_n(z^{(n)}|p), \tag{44}$$

say. The distribution function F of X_1 and the empirical one, F_n, are respectively determined by the vectors $p = (p_1, \cdots, p_k)$ and $Z^{(n)}$ whose values lie in the simplex

$$\Omega = \left\{(x_1, \cdots, x_k)|x_1 \geq 0, \cdots, x_k \geq 0, \sum_{i=1}^{k} x_i = 1\right\}. \tag{45}$$

It will be convenient to write $I(Z^{(n)}, p)$ for $I(F_n, F)$, where

$$I(x, p) = \sum_{i=1}^{k} x_i \log (x_i/p_i) \tag{46}$$

for x and p in Ω. We have

$$P_n(z^{(n)}|p) = P_n(z^{(n)}|z^{(n)}) \exp \{-nI(z^{(n)}, p)\}, \tag{47}$$

which corresponds to equations (35) and (43).

For any set $A \subset \Omega$, let $I(A, p) = \inf \{I(x, p)|x \in A\}$ and let $A^{(n)}$ denote the set of points $z^{(n)} \in A$. In [5] it is shown that

$$P\{Z^{(n)} \in A\} = \exp \{-nI(A^{(n)}, p) + O(\log n)\} \tag{48}$$

uniformly for $A \subset \Omega$ and $p \in \Omega$. Clearly, $I(A^{(n)}, p) \geq I(A, p)$. Hence, if $\{A_n\}$ is a sequence of sets such that

$$I(A_n^{(n)}, p) \leq I(A_n, p) + O(n^{-1} \log n), \tag{49}$$

then

$$P\{Z^{(n)} \in A_n\} \asymp n^{r_n} \exp \{-nI(A_n, p)\}, \tag{50}$$

where r_n is bounded. Sufficient conditions for (49) to hold are given in the appendix of [5].

Here we shall consider the determination of the order of magnitude of $P\{Z^{(n)} \in A_n\}$, which amounts to the determination of r_n in (50). The point p will be held fixed with $p_i > 0$ for all i. (The results to be derived hold uniformly for $p_i > \epsilon$, $i = 1, \cdots, k$, where ϵ is any fixed positive number.)

LEMMA 3. *For every real m there is a number d (which depends only on m and k) such that uniformly for $A \subset \Omega$,*

$$\begin{aligned} P\{Z^{(n)} \in A\} = {} & P\{Z^{(n)} \in A, I(Z^{(n)}, p) < I(A, p) + dn^{-1} \log n\} \\ & + O(n^{-m} \exp \{-nI(A, p)\}). \end{aligned} \tag{51}$$

The lemma follows from (48) with A replaced by $\{x|I(x, p) \geq I(A, p) + dn^{-1} \log n\}$ and d suitably chosen.

It should be noted that $Z^{(n)} \in A$ implies $I(Z^{(n)}, p) \geq I(A, p)$. Thus if the remainder term in (51) is negligible, the main contribution to $P\{Z^{(n)} \in A\}$ is from the intersection of A with a narrow strip surrounding the (convex) set $\{x|I(x, p) < I(A, p)\}$.

Let Ω_ϵ denote the set of all $x \in \Omega$ such that $x_i > \epsilon$, $i = 1, \cdots, k$. Let for $x \in \Omega$,

$$\Pi_n (x|p) = (2\pi n)^{-(k-1)/2} \left(\prod_{i=1}^{k} x_i\right)^{-1/2} \exp \{-nI(x, p)\}. \tag{52}$$

LEMMA 4. *For $\epsilon > 0$ fixed we have uniformly for $A \subset \Omega_\epsilon$,*

$$P\{Z^{(n)} \in A\} = \sum_{z^{(n)} \in A} \Pi_n(z^{(n)}|p)(1 + O(n^{-1})). \tag{53}$$

This follows from (47) by applying Stirling's formula to

$$P_n(z^{(n)}|z^{(n)}) = (n!/n^n) \prod_{i=1}^{k} (n_i^{n_i}/n_i!). \tag{54}$$

We now approximate the sum in (53) by an integral. To determine the order of magnitude only, a crude approximation will suffice. Let

$$\begin{aligned} R_n(z^{(n)}) &= \{(x_1, \cdots, x_{k-1})|z_i^{(n)} \le x_i < z_i^{(n)} + n^{-1}, i = 1, \cdots, k-1\}, \\ A_n^* &= \bigcup_{z^{(n)} \in A} R_n(z^{(n)}). \end{aligned} \tag{55}$$

LEMMA 5. *For $\epsilon > 0$ we have uniformly for $A \subset \Omega_\epsilon$,*

$$\sum_{z^{(n)} \in A} \Pi_n(z^{(n)}|p) \asymp n^{k-1} \int \underset{A_n^*}{\cdots} \int \Pi_n(x|p)\, dx_1 \cdots dx_{k-1}. \tag{56}$$

PROOF. We have $1 = n^{k-1} \int \cdots \int_{R_n(z^{(n)})} dx_1 \cdots dx_{k-1}$. If $(x_1, \cdots, x_{k-1}) \in R_n(z^{(n)})$, then $|x_i - z_i^{(n)}| < kn^{-1}$ for $i = 1, \cdots, k$, where $x_k = 1 - x_1 \cdots x_{k-1}$. Also, $I(x, p) = I(z, p) + 0(\max_i |x_i - z_i|)$ uniformly for x and $z \in \Omega_\epsilon$. These facts imply the lemma.

Now let $f(x)$ be a function defined on Ω,

$$A(c) = \{x|f(x) \ge c\}, \tag{57}$$

and suppose that for every $\epsilon' > 0$ there is a number $a_1(\epsilon')$ such that

$$|f(z) - f(x)| \le a_1(\epsilon')|z - x| \qquad \text{if} \quad z \in \Omega_{\epsilon'}, \quad x \in \Omega_{\epsilon'}, \tag{58}$$

where $|z - x| = \max_i |z_i - x_i|$. This condition is satisfied if the first partial derivatives of f exist and are continuous in Ω_0 (the set where $x_i > 0$ for all i). Let

$$D(c, \delta) = \{x|f(x) \ge c, \quad I(x, p) \le I(A(c), p) + \delta\}, \tag{59}$$

$$D^*(c, \delta) = \{(x_1, \cdots, x_{k-1})|(x_1, \cdots, x_{k-1}, 1 - x_1 - \cdots - x_{k-1}) \in D(c, \delta)\}, \tag{60}$$

$$V_c(u) = \int \cdots \int_{D^*(c,u)} dx_1 \cdots dx_{k-1}, \tag{61}$$

and, if the derivative $V_c'(u) = dV_c(u)/du$ exists for $0 < u < \delta$,

$$K_n(c, \delta) = \int_0^\delta e^{-nu} V_c'(u)\, du. \tag{62}$$

THEOREM 1. *Let $A(c)$ be defined by (57), where f satisfies (58) for every $\epsilon' > 0$. Let $\{c_n\}$ be a real number sequence and suppose that for every $a' > 0$ there are positive numbers ϵ, δ, and n_0 such that $D(c_n - a'n^{-1}, \delta) \subset \Omega_\epsilon$ for $n > n_0$. Then for every real number m there are positive numbers d and a such that*

$$\begin{aligned} &P\{f(Z^{(n)}) \ge c_n\} \\ &\quad \asymp \exp\{-nI(A(c_n), p)\}\{n^{(k-1)/2} K_n(c_n - \theta a n^{-1}, \delta_n + \theta a n^{-1}) + O(n^{-m})\}, \end{aligned} \tag{63}$$

where $|\theta| \le 1$, $\delta_n = dn^{-1} \log n$, and it is assumed that for each c such that $|c - c_n| \le an^{-1}$ the derivative $V_c'(u)$ exists for $0 < u < \delta$.

PROOF. From lemmas 3, 4, and 5 we obtain

$$(64)\qquad P\{f(Z^{(n)}) \geq c_n\} \asymp n^{(k-1)/2} J_{1,n} + 0(n^{-m} \exp\{-nI(A(c_n), p\}),$$

where

$$(65)\qquad J_{1,n} = \int \cdots \int_{D_n} \left(\prod_{i=1}^{k} x_i\right)^{-1/2} \exp\{-nI(x, p)\}\, dx_1 \cdots dx_{k-1},$$

$$(66)\qquad D_n^* = \bigcup_{z^{(n)} \in D(c_n, \delta_n)} R_n(z^{(n)}),$$

and $\delta_n = dn^{-1} \log n$. It follows from condition (58), which is also satisfied by $I(\cdot, p)$, that there is a number $a > 0$ such that

$$(67)\qquad D^*(c_n + an^{-1}, \delta_n - an^{-1}) \subset D_n^* \subset D^*(c_n - an^{-1}, \delta_n + an^{-1}).$$

Since $(\prod x_i)^{-1/2}$ is bounded in Ω_ϵ, we obtain $J_{1,n} \asymp J_{2,n}(c_n - \theta an^{-1}, \delta_n + \theta an^{-1})$, where

$$(68)\qquad J_{2,n}(c, \delta) = \int \cdots \int_{D^*(c,\delta)} \exp\{-nI(x, p)\}\, dx_1 \cdots dx_{k-1}$$

and $|\theta| \leq 1$. If the derivative $V_c'(u)$ exists for $0 < u < \delta$, we can write

$$(69)\qquad J_{2,n}(c, \delta) = \exp\{-nI(A(c), p)\} \int_0^\delta e^{-nu} V_c'(u)\, du.$$

The theorem follows.

If we had not suppressed the factor $(\prod x_i)^{-1/2}$, we would have obtained (63) with $V_c(u)$ replaced by

$$(70)\qquad V_{1,c}(u) = \int \cdots \int_{D^*(c,u)} \left\{\prod_{i=1}^{k} x_i\right\}^{-1/2} dx_1 \cdots dx_{k-1}.$$

In this form the first term on the right of (63) is analogous to the right side of (37). The integer k in (37) is here replaced by $k - 1$, since the distribution is $(k - 1)$-dimensional.

To apply theorem 1 we need to determine the order of magnitude of $K_n(c_n, \delta_n)$, where $\delta_n \asymp n^{-1} \log n$, so that $\delta_n \to 0$ and $n\delta_n \to \infty$. If, for instance $V_{c_n}'(u) \asymp b(c_n)u^r$ uniformly with respect to n as $u \to 0^+$, then

$$(71)\qquad K_n(c_n, \delta_n) \asymp b(c_n) \int_0^{\delta_n} e^{-nu} u^r\, du \asymp b(c_n) n^{-r-1}.$$

Concerning the determination of the order of magnitude of $V_c'(u)$, we observe the following. Note that $I(A(c), p) = 0$ if $f(p) \geq c$. Assume that $f(p) < c$. The continuity condition (58) implies that $I(A(c), p) > 0$. Let Y denote the set of points $y \in A(c)$ such that $I(y, p) = I(A(c), p)$. Then $Y \subset D(c, \delta)$. The assumption $D(c, \delta) \subset \Omega_\epsilon$, condition (58), and the convexity of $I(\cdot, p)$ imply that $f(y) = c$ if $y \in Y$.

Suppose first that the set $A(c)$ is contained in a half-space H such that $I(H, p) = I(A(c), p)$. (This is true if the function $-f(x)$ is convex, so that the set $A(c)$ is convex.) Then the set Y consists of a single point y, and we have

$$(72)\qquad \begin{aligned} I(x, p) - I(A(c), p) &= I(x, p) - I(y, p) \\ &= \sum (\log (y_i/p_i))(x_i - y_i) + I(x, y). \end{aligned}$$

Hence, $H = \{x \mid \sum (\log y_i/p_i)(x_i - y_i) \geq 0\}$, and $x \in A(c)$ implies $I(x, p) - I(A(c), p) \geq I(x, y)$. Therefore, if $x \in D(c, \delta_n)$, then $I(x, y) < \delta_n$. Now $I(x, y) = \frac{1}{2}Q^2(x, y) + O(|x - y|^3)$, where $Q^2(x, y) = \sum (x_i - y_i)^2/y_i$. Hence, if $x \in D(c, \delta_n)$, then $|x - y| = O(\delta_n^{1/2})$, $I(x, y) - \frac{1}{2}Q^2(x, y) = 0(\delta_n^{3/2}) = o(n^{-1})$, and the inequality

$$I(x, p) - I(A(c), p) < \delta_n \tag{73}$$

may be written

$$\sum (\log (y_i/p_i))(x_i - y_i) + \tfrac{1}{2}Q^2(x, y) < \delta_n + o(n^{-1}). \tag{74}$$

An inspection of the proof of theorem 1 shows that in the present case the theorem remains true if in the domain of integration $D(c, u)$ of the integral $V_c(u)$, the left-hand side of (73) is replaced by the left-hand side of (74).

Now suppose further that the partial derivatives $f'_i(x) = \partial f(x)/\partial x_i$, $f''_{ij}(x) = \partial^2 f(x)/\partial x_i \partial x_j$, and the third-order derivatives exist and are continuous in Ω_0. Then

$$f(x) - c = f(x) - f(y) = \sum f'_i(y)(x_i - y_i) + \tfrac{1}{2}F(x - y) + O(|x - y|^3), \tag{75}$$

uniformly for $y \in \Omega_\epsilon$, where $F(x - y) = \sum \sum f''_{ij}(y)(x_i - y_i)(x_j - y_j)$. Hence, if $x \in D(c, \delta_n)$, the inequality $f(x) \geq c$ may be written as

$$\sum f'_i(y)(x_i - y_i) + \tfrac{1}{2}F(x - y) \geq r_n, \qquad r_n = O(\delta_n^{3/2}). \tag{76}$$

Furthermore, the half-space $\{x \mid \sum f'_i(y)(x_i - y_i) \geq 0\}$ is identical with H. This implies that $y = y(c)$ satisfies the equations

$$\log (y_i/p_i) = t(c)f'_i(y) + s(c), \qquad i = 1, \cdots, k, \tag{77}$$

where $t(c) > 0$ and $s(c)$ are constants, as well as the equations $f(y) = c$ and $\sum y_i = 1$. It follows that under the present assumptions theorem 1 remains true with $K_n(c_n, \delta_n)$ replaced by $K_n(c_n, \delta_n, r_n)$, where

$$K_n(c, \delta, r) = \int_0^\delta e^{-nu} V'_{c,r}(u)\, du \tag{78}$$

and $V'_{c,r}(u)$ is the derivative with respect to u of the volume $V_{c,r}(u)$ of $D^*(c, r, u)$, the set of points $(x_1, \cdots, x_{k-1})$ which satisfy the inequalities

$$\sum f'_i(y)(x_i - y_i) + \tfrac{1}{2}F(x - y) > r, \tag{79}$$

$$t(c) \sum f'_i(y)(x_i - y_i) + \tfrac{1}{2}Q^2(x, y) < u. \tag{80}$$

If we make the substitution $z_i = y_i^{-1/2}(x_i - y_i)$, $i = 1, \cdots, k$, we obtain $Q^2(x, y) = \sum_{i=1}^k z_i^2$, $\sum y_i^{1/2} z_i = 0$ and

$$\sum f'_i(y)(x_i - y_i) = \sum (f'_i(y) - a(c))y_i^{1/2} z_i = \sigma(c) \sum b_i z_i, \tag{81}$$

where $a(c) = \sum y_i f'_i(y)$, $\sigma^2(c) = \sum (f'_i(y) - a(c))^2 y_i$, and $b_i = \sigma^{-1}(c)(f'_i(y) - a(c))$. We have $\sum b_i^2 = 1$, $\sum b_i y_i^{1/2} = 0$. Hence we can perform an orthogonal transformation $(z_1, \cdots, z_k) \to (v_1, \cdots, v_k)$, where $v_1 = \sum b_i z_i = \sigma^{-1}(c) \sum f'_i(y_i)(x_i - y_i)$ and $v_k = \sum y_i^{1/2} z_i = 0$. The inequalities (79), (80) are transformed into

$$\sigma(c)\, v_1 + \tfrac{1}{2}G(v_1, \cdots, v_{k-1}) > r, \tag{82}$$

$$t(c)\sigma(c)v_1 + \tfrac{1}{2}\sum_{i=1}^{k-1} v_i^2 < u, \tag{83}$$

where $G(v_1, \cdots, v_{k-1})$ is a quadratic form in $v_1, \cdots, v_{k-1}$. Thus

$$V_{c,r}(u) = C(c)W_{c,r}(u), \qquad V'_{c,r}(u) = C(c)W'_{c,r}(u), \tag{84}$$

where $C(c)$ is the modulus of the determinant of the linear transformation $(x_1, \cdots, x_{k-1}) \to (v_1, \cdots, v_{k-1})$ and $W_{c,r}(u)$ is the volume of the set defined by (82) and (83). In the estimation of $W'_{c,r}(u)$ we may assume that $u = 0(\delta_n)$ and $r = O(\delta_n^{3/2}) = o(n^{-1})$.

The replacement of $V_c(u)$ by $V_{c,r}(u)$ may be possible under conditions different from those assumed in the two preceding paragraphs. Suppose that $c > f(p)$ is fixed and that the set Y consists of a finite number s of points. Choose $\eta > 0$ so small that the s sets $S_y = \{x| \, |x - y| < \eta\}$, $y \in Y$, are disjoint. Then for δ small enough $D(c, \delta)$ is contained in the union of the sets S_y. If, for each $y \in Y$, the surfaces $f(x) = c$ and $I(x, y) = I(A(c), p)$ are not too close in the neighborhood of y, then $x \in S_y \cap D(c, \delta)$ will imply $|x - y| = 0(\delta^{1/2})$, and we arrive at analogous conclusions as in the preceding case.

If $f(x) = \sum_{i=1}^k a_i x_i$ is a linear function, $nf(z^{(n)})$ is the sum of n independent random variables, each of which takes the values $a_1, \cdots, a_k$ with respective probabilities $p_1, \cdots, p_k$. The following theorem can be deduced from theorem 1 but is a special case of (13).

THEOREM 2. *Let $A(c) = \{x| \sum_{i=1}^k a_i x_i \geq c\}$, where $a_1, \cdots, a_k$ are fixed (not all equal) and $\sum_{i=1}^k a_i p_i = 0$. Then*

$$P\left\{\sum_{i=1}^{k} a_i Z_i^{(n)} \geq c\right\} \asymp c^{-1} n^{-1/2} \exp\{-nI(A(c), p)\} \tag{85}$$

uniformly for $\alpha n^{-1/2} < c < \max a_i - \beta$, where α and β are arbitrary positive constants.

The next theorem gives an analogous uniform estimate for the distribution of $I(Z^{(n)}, p)$.

THEOREM 3. *Let $p_{\min} = \min_i p_i$ and let α and β be arbitrary positive constants. Then*

$$P\{I(Z^{(n)}, p) \geq c\} \asymp (nc)^{(k-3)/2} e^{-nc} \tag{86}$$

uniformly for $\alpha n^{-1} < c < -\log(1 - p_{\min}) - \beta$.

PROOF. In this case $D(c, \delta) = \{x | c \leq I(x, p) < c + \delta\}$. It can be shown that $I(x, p) < -\log(1 - p_{\min})$ implies $x_i > 0$ for all i. The assumption $c < -\log(1 - p_{\min}) - \beta$, $\beta > 0$, implies $D(c, \delta) \subset \Omega_\epsilon$ for some $\epsilon > 0$ if δ is small enough. Let

$$V(u) = \int \cdots \int_{I(x,p) < u} dx_1 \cdots dx_{k-1}. \tag{87}$$

We first prove the following lemma. (Clearly we may assume $p_k = p_{\min}$.)

LEMMA 6. *Let $p_k = p_{\min}$. The derivative $V'(u) = dV(u)/du$ exists, is continuous, and positive for $0 < u < -\log(1 - p_{\min})$ and $V'(u) \asymp u^{(k-3)/2}$ as $u \to 0^+$.*

(Heuristically, as $u \to 0$, $V(u)$ is approximated by the volume of the ellipsoid $Q^2(x, p) < 2u$, which is proportional to $u^{(k-1)/2}$.)

We shall write $I_k(x, p)$ for $I(x, p)$ to indicate the number of components of the arguments x and p, and $V_{k,p}(u)$ for $V(u)$. For $x_k \neq 1$, let $y = (y_1, \cdots, y_{k-1})$, $y_i = x_i/(1 - x_k)$; $z = (z_1, z_2) = (1 - x_k, x_k)$; $q = (q_1, \cdots, q_{k-1})$, $q_i = p_i/(1 - p_k)$; $r = (r_1, r_2) = (1 - p_k, p_k)$. Then we have the identity

$$I_k(x, p) = z_1 I_{k-1}(y, q) + I_2(z, r). \tag{88}$$

Hence, we obtain the recurrence relation

$$V_{k,p}(u) = \int_{I_2(z,r)<u} z_1^{k-2} V_{k-1,q}(z_1^{-1}\{u - I_2(z, r)\})\, dz_1, \qquad k \geq 3. \tag{89}$$

Since $p_k = p_{\min} \leq 1 - p_k$, we have $r_1 \geq r_2$. We also have $V_{2,r}(u) = b(u) - a(u)$ where, for $0 < u < -\log(1 - p_{\min}) = -\log r_2$, $a(u)$ and $b(u)$ are the two roots of the equation $I_2((z_1, 1 - z_1), r) = u$, $0 < a(u) < r_1 < b(u) < 1$. Hence, it is easy to show that the lemma is true for $k = 2$. From (89) we obtain for $k = 3$,

$$V'_{k,p}(u) = \int_{a(u)}^{b(u)} z_1^{k-3} V'_{k-1,q}(z_1^{-1}\{u - I_2(z, r)\})\, dz_1. \tag{90}$$

It now can be shown that the lemma holds for $k = 3$ and, by induction, that equation (90) and the lemma are true for any k.

Under the conditions of theorem 1 we have $V_c(u) = V(c + u) - V(c)$. The lemma implies that $V'_c(u) = V'(c + u) \asymp (c + u)^{(k-3)/2}$ uniformly for $0 < c + u < -\log(1 - p_{\min}) - \beta$. It follows that uniformly for $\alpha n^{-1} < c < -\log(1 - p_{\min}) - \beta$,

$$K_n(c, \delta_n) \asymp \int_0^{\delta_n} e^{-nu}(c + u)^{(k-3)/2}\, du \asymp c^{(k-3)/2} n^{-1}. \tag{91}$$

This establishes the theorem under the restriction $nc > a + \alpha$, where a is the number which appears in (63). That the result holds for $nc > \alpha$ with any $\alpha > 0$ follows from the well-known fact that $2nI(Z^{(n)}, p)$ has a chi-square limit distribution.

Since $A \subset \{x | I(x, p) \geq I(A, p)\}$ for any subset A of Ω, theorem 3 immediately implies the following theorem.

THEOREM 4. *If α and β are any positive numbers, there is a constant $C = C(\alpha, \beta, p)$ such that for any set A which satisfies*

$$\alpha n^{-1} < I(A, p) < -\log(1 - p_{\min}) - \beta, \tag{92}$$

we have

$$P\{Z^{(n)} \in A\} \leq C\{nI(A, p)\}^{(k-3)/2} \exp\{-nI(A, p)\}. \tag{93}$$

REMARK. We have $\max_{x \in \Omega} I(x, p) = -\log p_{\min}$. It seems plausible that the estimate (86) of theorem 3 holds uniformly for $\alpha n^{-1} < c < -\log p_{\min} - \beta$. If so, theorem 4 holds with an analogous modification.

For the functions $f(x) = \sum a_i x_i$ and $f(x) = I(x, p)$ of theorems 2 and 3 the order of magnitude of $P\{f(Z^{(n)}) \geq c\}$ is expressed in the form $c^r n^s$

$\exp\{-nI(A(c), p)\}$ in a wide range of c. That this is not true in general is shown by the following example. Let

$$A(c) = \{x|Q^2(x, p) \geq c\}, \qquad Q^2(x, p) = \sum (x_i - p_i)^2/p_i. \tag{94}$$

Since $I(x, p) = \frac{1}{2}Q^2(x, p) + 0(|x - p|^3)$, $Q^2(x, p) < c$ implies $I(x, p) = \frac{1}{2}Q^2(x, p) + O(c^{3/2})$ as $c \to 0$. Hence, it follows from theorem 3 that if $c = O(n^{-2/3})$ and $nc > \alpha > 0$, then $P\{Q^2(Z^{(n)}, p) \geq c\}$ is of the same order of magnitude as $P\{I(Z^{(n)}, p) \geq c/2\}$. We have $I(A(c), p) = \frac{1}{2}c + O(c^{3/2})$. By theorem 3 this implies that

$$P\{Q^2(Z^{(n)}, p) \geq c\} \asymp (nc)^{(k-3)/2} \exp\{-nI(A(c), p)\} \tag{95}$$

uniformly for $\alpha n^{-1} < c < \beta n^{-2/3}$. On the other hand, if c is bounded away from 0 and from $\max_x Q^2(x, p) = p_{\min}^{-1} - 1$, it can be deduced from theorem 1 (see the remarks after the proof of theorem 1 and section 8 of [5]) that

$$P\{Q^2(Z^{(n)}, p) \geq c\} \asymp n^{-1/2} \exp\{-nI(A(c), p)\}. \tag{96}$$

In this case the probability of the set $A(c)$ is of the same order of magnitude as the probability of any of the half-spaces contained in $B = \{x|I(k, p) \geq I(A(c), p)\}$ and bounded by the supporting hyperplanes of the convex set $B' = \{x|I(x, p) < I(A(c), p)\}$ at the common boundary points y of the sets $A(c)$ and B. This result holds for a wide class of functions f when c is fixed. (However, theorem 4 of [12] is inaccurate in the stated generality, as is seen from theorem 3 above.) An asymptotic expression for $P\{Q^2(Z^{(n)}, p) \geq c\}$ with $c = o(1)$ as $n \to \infty$ has been obtained by Richter [11].

REFERENCES

[1] R. R. Bahadur and R. Ranga Rao, "On deviations of the sample mean," *Ann. Math. Statist.*, Vol. 31 (1960), pp. 1015–1027.

[2] A. A. Borovkov and B. A. Rogozin, "On the central limit theorem in the multidimensional case," *Teor. Verojatnost. i Primenen.*, Vol. 10 (1965), pp. 61–69. (In Russian.)

[3] H. Cramér, "Sur un nouveau théorème limite de la théorie des probabilités," *Actualités Sci. Indust.*, No. 736 (1938).

[4] W. Feller, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.

[5] Wassily Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 369–401.

[6] J. Kiefer and J. Wolfowitz, "On the deviations of the empiric distribution function of vector chance variables," *Trans. Amer. Math. Soc.*, Vol. 87 (1958), pp. 173–186.

[7] Yu. V. Linnik, "Limit theorems for sums of independent variables taking into account large deviations, I, II, III," *Teor. Verojatnost. i Primenen.*, Vol. 6 (1961), pp. 145–163 and 377–391; Vol. 7 (1962), pp. 122–134. (In Russian.) (English translation in *Theor. Probability Appl.*, Vol. 6 (1961), pp. 131–148 and 345–360; Vol. 7 (1962), pp. 115–129.)

[8] V. V. Petrov, "Generalization of Cramér's limit theorem," *Uspehi Mat. Nauk* (n.s.), Vol. 9 (1954), pp. 195–202. (In Russian.)

[9] ———, "Limit theorems for large deviations when Cramér's condition is violated, I, II," *Vestnik Leningrad Univ. Ser. Mat. Meh. Astronom.*, Vol. 18 (1963), pp. 49–68; Vol. 19 (1964), pp. 58–75. (In Russian.)

[10] R. RANGA RAO, "On the central limit theorem in R_k," *Bull. Amer. Math. Soc.*, Vol. 67 (1961), pp. 359–361.

[11] W. RICHTER, "Mehrdimensionale Grenzwertsätze für grosse Abweichungen und ihre Anwendung auf die Verteilung von χ^2," *Teor. Verojatnost. i Primenen.*, Vol. 9 (1964), pp. 31–42.

[12] I. N. SANOV, "On the probability of large deviations of random variables," *Mat. Sbornik* (n.s.), Vol. 42 (1957), pp. 11–44. (In Russian.) (English translation in *Select. Transl. Math. Statist. and Probability*, Vol. 1 (1961), pp. 213–244.)

[13] J. SETHURAMAN, "On the probability of large deviations of families of sample means," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 1304–1316.

[14] YU. V. LINNIK, "On the probability of large deviations for the sums of independent variables," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1961, Vol. 2, pp. 289–306.

[15] A. V. NAGAEV, "Large deviations for a class of distributions," *Predel'nye Teoremy Teorii Verojatnostei* (1963), pp. 56–58, Akad. Nauk Uzb. SSR, Tashkent. (In Russian.)

[16] S. V. NAGAEV, "An integral limit theorem for large deviations," *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk*, Vol. 6 (1962), pp. 37–43. (In Russian.)