Pacific Journal of Mathematics

ON LARGE DEVIATIONS OF THE EMPIRIC D. F. OF VECTOR CHANCE VARIABLES AND A LAW OF THE ITERATED LOGARITHM

JACK CARL KIEFER

Vol. 11, No. 2 December 1961

ON LARGE DEVIATIONS OF THE EMPIRIC D.F. OF VECTOR CHANCE VARIABLES AND A LAW OF THE ITERATED LOGARITHM

J. KIEFER

1. Introduction and preliminaries. Let F be a distribution function (d.f.) on m-dimensional Euclidean space R^m, and let $X_1, \cdots, X_n$ be independent chance vectors with common d.f. F. The empiric d.f. S_n is a chance d.f. on R^m defined as follows: if $x = (x_1, \cdots, x_m)$, $nS_n(x)$ is the number of X_i's, $1 \leqq i \leqq n$, such that, for $j = 1, \cdots, m$, the jth component $X_i^{(j)}$ of X_i is less than or equal to x_i.

When $m = 1$, the distribution of $D_n = \sup_x |S_n(x) - F(x)|$ is the same for all continuous F, and Kolomogorov [5] first computed the limiting distribution of $n^{1/2}D_n$ as $n \to \infty$. Chung [1] gave a bound on the error term which was sharp enough to yield a law of the iterated logarithm for the empiric d.f. and, in fact, the more precise complete characterization of monotone functions of upper and lower class. (The more recent literature contains several asymptotic expansions of Kolmogorov's distribution.) It was proved by Dvoretzky, Kiefer and Wolfowitz [2] that there is a universal constant C such that, for all $n > 0$ and $r \geqq 0$,

$$P\{n^{1/2}D_n \geqq r\} \leqq Ce^{-2r^2}; \tag{1.1}$$

since $\lim_n P\{n^{1/2}D_n \geqq r\}$ is asymptotically $2e^{-2r^2}(1 + o(1))$ as $r \to \infty$, the estimate (1.1) cannot be improved upon in this general form.

Much less is known when $m > 1$. The limiting d.f. of $n^{1/2}D_n$ was proved to exist by Kiefer and Wolfowitz [4]; of course, its form depends on F (and is unknown except in a few trivial cases), unlike the case $m = 1$. It was also proved in [4] that there exist positive constants c_m and c'_m such that, for all F, $n > 0$, and $r \geqq 0$,

$$P_F\{n^{1/2}D_n \geqq r\} \leqq c'_m e^{-c_m r^2}, \tag{1.2}$$

whereby $P_F\{A\}$ we denote the probability of the event A when X_1 has d.f. F. Possible choices of the constants c_m were shown to be

$$c_2 = .0157,\ c_3 = .000107, \cdots (\lim_m c_m = 0). \tag{1.3}$$

It was also shown in [4] that, for $m > 1$, one cannot have $c_m = 2$ in (1.2); specifically, if $m = 2$ and F^* distributes probability uniformly on the line segment $\{(x_1, x_2): x_1 \geqq 0, x_2 \geqq 0, x_1 + x_2 = 1\}$, then as $r \to \infty$ we have

$$\lim_n P_{F^*}\{n^{1/2}D_n \geqq r\} = 8r^2e^{-2r^2}(1 + o(1)). \tag{1.4}$$

Received May 9, 1960. Research sponsored by the Office of Naval Research.

Thus, even for a single F we cannot hope to achieve (1.1) in the case $m > 1$.

The main object of the present paper is to prove the next best thing, namely,

THEOREM 1-m. *For each m and $\varepsilon > 0$ there is a constant $c(\varepsilon, m)$ such that, for all F, $n > 0$, and $r \geqq 0$, we have*

$$P_F\{n^{1/2}D_n \geqq r\} \leqq c(\varepsilon, m)e^{-(2-\varepsilon)r^2} . \tag{1.5}$$

(The labeling of the Theorem as 1-m is to make clear the induction on m in the proof.)

The result (1.5) clearly represents a marked improvement over (1.3), and in view of (1.4) this result (1.5) is the best possible of this form. Whether or not the ε in the exponent can be improved, e.g., to a term like $cr^{-2} \log r$ when $m = 2$, as in (1.4), is not known, and the methods of the present paper do not seem capable of shedding any light on the subject.

The weaker result which is obtained by replacing the left side of (1.5) by its limit as $n \to \infty$ can be proved more quickly, and the reader will have no difficulty in recognizing how the proof of the present paper can be shortened if only that result on the corresponding limiting Gaussian process (see [4]) is desired.

Theorem 1 yields an easy proof of a law of the iterated logarithm:

THEOREM 2. *For every m and every continuous F,*

$$P_F\{\limsup_{n\to\infty} n^{1/2}D_n / (2^{-1} \log \log n)^{1/2} = 1\} = 1 . \tag{1.6}$$

(*The same result holds for $D_n^+ = \sup_x [S_n(x) - F(x)]$ or for $D_n^- = \sup_x [F(x) - S_n(x)]$.*)

In fact, the conclusion that the upper limit is $\geqq 1$ follows at once on applying Chung's result to a one-dimensional marginal d.f. and empiric d.f. On the other hand, the proof that, if $\lambda > 2^{-1}$, the probability is zero that $n^{1/2}D_n \geqq (\lambda \log \log n)^{1/2}$ for infinitely many n, is proved in a classical way. For example, the proof on page 48 of [1] requires only trivial modifications to apply to the present case, using the estimate (1.5) for ε sufficiently small and the fact that for some positive constants b and c we have $P_F\{D_n \leqq bn^{-1/2}\} \geqq c > 0$ for all n; the latter is an immediate consequence of Theorem 1-m or of the results of [4]. It is unnecessary to give more details of the proof of Theorem 2.

Obviously, our estimate (1.5) is not precise enough to yield a finer result analogous to Chung's. In fact, it is clear that the value of λ which divides the functions $[2^{-1}\log \log n + \lambda \log \log \log n]^{1/2}$ into upper and lower classes depends on F. For example, when $m = 2$ and $F = F^*$,

the d.f. of $n^{1/2}D_n$ is that of $n^{1/2}(D_n^+ + D_n^-)$ for $m = 1$ (see [4]) and, following Chung in obtaining an error term in the approach to the limiting distribution and in the characterization of the upper and lower classes, we obtain 3/2 for the critical value of λ. On the other hand, if F distributes probability uniformly on the main diagonal $x_1 = x_2$ of the unit square, D_n has the same distribution as in the case $m = 1$, and Chung's result then yields $\lambda = 1$ as the critical value.

The method used to prove Theorem 1-m is an improvement of the method used to prove Theorem 1 of [4], the line of argument being similar. Lemma 2 extends equation (2.4) of [4], where the case $k = 2$, $j = 1$ was considered. Lemma 3-m improves the estimate of equation (2.5) of [4]. Lemma 1 is needed in order to obtain, in Lemma 6-m, an improvement on the estimate of equation (2.24) of [4]. However, the present paper is self-contained, and we shall not make use of the results of [2] or [4] in the proof.

The idea of using an argument like that of Lemma 2 is well known in such a context as that of the study of the maximum of partial sums of independent summands, where it is of course much easier to apply than in the case of "tieddown" processes such as the random functions $S_n(x) - F(x)$ studied here. In fact, it is just as easy to obtain such results for processes with independent increments in the case of m-dimensional "time" as it is in the classical case of one-dimensional time. For example, if $X_{i_1,i_2,\cdots,i_m}$, $1 \leqq i_j \leqq n_j$, $1 \leqq j \leqq m$, are independent random variables which are symmetric about 0 (these assumptions are easily relaxed) and

$$S_{i_1,i_2,\cdots,i_m} = \sum_{\substack{t_j \leqq i_j \\ 1 \leqq j \leqq m}} X_{t_1,t_2,\cdots,t_m},$$

we obtain, on using the classical argument m times,

$$P\{\max_{\substack{1 \leqq i_j \leqq n_j \\ 1 \leqq j \leqq m}} S_{i_1,\cdots,i_m} \geqq r\} \leqq 2^m P\{S_{n_1,\cdots,n_m} \geqq r\}.$$

Similary, the standard semi-martingale inequality on $E\{I_A U_n\}$, where $U_n = \max_{i \leqq n} Y_i$ (with $\{Y_i\}$ a semi-martingale), $A = \{U_n \geqq r\}$, and I_A is the indicator of A, has an obvious analogue for semi-martingales with m-dimensional time. Such results also carry over to the case of m-dimensional continuous time; for example, if $Y(t_1, \cdots, t_m)$ is a separable Gaussian process (m-dimensional time) with mean 0, independent stationary increments for t in the positive orthant, and $Y(t) = 0$ if t is outside the positive orthant, we have

$$P\{\sup_{\substack{0 \leqq t_i \leqq T_i \\ 1 \leqq i \leqq m}} Y(t_1, \cdots, t_m) \geqq r\} \leqq 2^m P\{Y(T_1, \cdots, T_m) \geqq r\}.$$

The results of the present paper yield inequalities for other processes with m-dimensional time, which require somewhat greater effort.

In the proof of Theorem 1-m we require one rather elementary result on the tails of the binomial distribution. Let $X_1, X_2, \cdots$ be independent random variables with $P_p\{X_1 = 1\} = 1 - P_p\{X_1 = 0\} = p$, where $0 < p < 1$ and we now use the subscript p to designate the underlying probability law. Write $\bar{X}_n = \sum_{i \leqq n} X_i/n$. It is well known that, for some $C > 0$,

$$P_p\{n^{1/2} \mid \bar{X}_n - p \mid \geqq r\} < Ce^{-2r^2} \tag{1.7}$$

for all $n > 1$, $r \geqq 0$, and p. In fact, (1.7) is much weaker than (1.1). The central limit theorem implies that, for p and r fixed, the limit of the left side of (1.7) as $n \to \infty$ is less than $C \exp\{-r^2/p(1-p)\}$, but it is well known (see the next paragraph or [6], p. 285) that the 2 in the exponent on the right side of (1.7) cannot be replaced by $c'/p(1-p)$ for all n, p, and r, where $c' > 0$. What we require is that the right side of (1.7) can be replaced by $C_p \exp\{-r^2 g(p)\}$, where $g(p) \to \infty$ as $p \to 0$.

We shall prove this inequality with $g(p) = \log(p^{-1}e^{-1})$ for $p < e^{-1}$. The factor e^{-1} in the expression for $g(p)$ can be improved slightly. However, the result cannot be improved by very much, since for $r = n^{1/2}(1-p)$ we obtain $\exp\{-r^2(\log p^{-1})(1-p)^{-2}\}$ for the left side of (1.7); thus, $g(p)$ cannot be taken to be $\varepsilon + \log p^{-1}$ with $\varepsilon > 0$. If $g(p)$ is allowed to depend upon r, it is easy to obtain a better bound. This is also true if one desires separate inequalities for the positive and negative deviations. But for $r \sim n^{1/2}(1-p)$ and $p \to 0$, it is again true that little improvement over our bound is possible, for the positive deviations.

LEMMA 1. *For $0 < p < e^{-1}$, all $n > 0$, and all $r \geqq 0$,*

$$P_p\{n^{1/2} \mid \bar{X}_n - p \mid \geqq r\} < C_p \exp\{-r^2[\log p^{-1} - 1]\}, \tag{1.8}$$

where C_p depends only on p.

Proof. We proceed along classical lines. In fact, for the negative deviations we need only note, putting $p' = 1 - p$, that

$$P_p\{n^{1/2}(\bar{X}_n - p) \leqq -r\} = P_{p'}\{n^{1/2}(\bar{X}_n - p') \geqq r\}, \tag{1.9}$$

and from [3], equations (VI. 10.12) and (VI. 3.5), we obtain easily that this last probability is bounded by

$$\begin{aligned} &C_p' \exp\{-r^2/2(1-p')\} = C_p' \exp\{-r^2/2p\} \\ &\quad \leqq C_p' \exp\{-r^2[\log p^{-1} - 1]\}, \end{aligned} \tag{1.10}$$

where here and in the sequel C_p, C_p', and C_p'' depend only on p but may change meaning in different appearances. We need a slightly better

bound than that just cited, for the positive deviations. Writing $b(K; n, p) = P_p\{n\bar{X}_n = k\}$, and letting the integer h be defined by $(n+1)p - 1 < h \leqq (n+1)p$, we have, for $k > h$,

$$\log \frac{b(k; n, p)}{b(h; n, p)} = \sum_{j=h+1}^{k} \log \frac{b(j; n, p)}{b(j-1; n, p)} = \sum_{j=h+1}^{k} \log \frac{p(n-j+1)}{(1-p)j} . \tag{1.11}$$

Since the summand is decreasing in j, we may bound the sum from above by integrating directly, with respect to j, from h to k. On combining terms, we obtain, for $n \geqq C_p'$ (which is all we need consider),

$$\log \frac{b(k; n, p)}{b(h; n, p)} < (k-h) \log \frac{p(n+1-h)}{(1-p)h} \tag{1.12}$$

$$+ (n+1-k) \log \frac{n+1-h}{n+1-k} - k \log \frac{k}{h} < C_p'' + (k-h) - k \log \frac{k}{h} .$$

Consider the function

$$g(x) = x - (x + p_1) \log \left(1 + \frac{x}{p_1}\right) + \frac{\log (p_1^{-1} e^{-1+p_1})}{(1-p_1)^2} x^2 \tag{1.13}$$

on the interval $[0, 1 - p_1]$. We see easily that $g(0) = g(1-p_1) = 0$, that g is concave and decreasing for

$$0 \leqq x \leqq [(1-p_1)^2/2 \log (p_1^{-1} e^{-1+p_1})] - p_1 ,$$

and that g is convex beyond this last point. We conclude that $g(x) \leqq 0$ on $[0, 1-p_1]$. Putting $x = (k-h)/n$ and $p_1 = h/n$, we conclude that the expression (1.12) is no greater than

$$C_p'' - \frac{[k-h]^2 \log (p_1^{-1} e^{-1})}{n} \tag{1.14}$$

for $h < k \leqq n$. Finally, since $|p_1 - p| < n^{-1}$, equation (1.14) yields

$$\log \frac{b(k; n, p)}{b(h; n, p)} < C_p'' - \frac{[k-h]^2 \log (p^{-1} e^{-1})}{n} . \tag{1.15}$$

Equations (1.10) and (1.15), together with the well known estimate $\sum_{j \geqq 0} b(k+j; n, p) < b(k; n, p)(k+1)(1-p)/[k+1-(n+1)p]$ relating tail probabilities to individual terms ([5], equation (VI. 3.5)), immediately yield (1.8).

2. Proof of Theorem 1-*m*. As indicated in [4], for any discontinuous d.f. F there is a continuous $\bar{F}$ for which the left side of (1.5) is at least as large. Hence, we can and do assume that F *is continuous in all that follows.* We can then transform the coordinates one at a

time, without changing the d.f. of D_n, so that *the marginal (one-dimensional)* d.f. *of* $X_1^{(j)}, 1 \leqq j \leqq m$, *is uniform on the interval* [0, 1]; we hereafter assume F to be of this form. We denote the unit m-cube, in which all X_i thus fall with probability one (w.p.1), by I^m.

In all that follows we need only consider values r for which $R(\varepsilon, m) < r \leqq n^{1/2}$, where $R(\varepsilon, m)$ is a fixed positive number. For $P_F\{n^{1/2}D_n > n^{1/2}\} = 0$, and if (1.5) is true for $r > R(\varepsilon, m)$ we can increase $c(\varepsilon, m)$ to make it hold for all $r \geqq 0$.

Let k be a real number $\geqq 2$ and let j be an integer, $1 \leqq j \leqq k-1$. Define

$$V_{jk} = \{(x_1, \cdots, x_m) : \ (j-1)/k \leqq x_1 \leqq j/k\} ,$$
$$W_{jk} = \{(x_1, \cdots, x_m) : \ x_1 = j/k\} .$$

We first show that if $S_n(x) - F(x)$ is $< rn^{-1/2}$ everywhere on $W_{j-1,k}$ and $\geqq rn^{-1/2}$ somewhere on the slab V_{jk}, then there is appreciable conditional probability that it is almost this large somewhere on the hyperplane W_{jk}. To this end, define the events

$$B_{jk}(r) = \left\{\sup_{V_{jk}} [S_n(x) - F(x)] \geqq rn^{-1/2}\right\} , \tag{2.1}$$
$$C_{jk}(r) = \left\{\sup_{W_{jk}} [S_n(x) - F(x)] \geqq rn^{-1/2}\right\} .$$

Denote the complement of an event C by $\bar{C}$.

LEMMA 2. *We have, for* $2 \leqq r \leqq n^{1/2}$, $k \geqq 2$, *and* $1 \leqq j \leqq k-1$,

$$P_F\left\{C_{jk}\left(r\left(1 - \frac{2}{k-j}\right)\right) \mid \bar{C}_{j-1,k}(r) B_{jk}(r)\right\} \tag{2.2}$$
$$\geqq 1 - \frac{(k-j)^2}{r^2} .$$

Proof. If $B_{jk}(r)$ occurs, there is w.p.1 a smallest value x_1' of x_1 ($j - 1 \leqq kx_1' \leqq j$) for which $S_n(x) - F(x) \geqq rn^{-1/2}$, for some $x_2, \cdots, x_m$; for $x_1 = x_1'$, there is then a smallest value x_2' of x_2 for which this inequality is satisfied, and so on. Thus, we obtain a well-defined random vector $X' = (x_1', \cdots, x_m')$ for which $S_n(X') - F(X') \geqq rn^{-1/2}$ whenever $B_{jk}(r)$ occurs. Moreover, the event $\{x_1' = a_1, \cdots, x_m' = a_m\}$ depends only on $\{S_n(x), x'_1 \leqq a_1\}$. When $a = (a_1, \cdots, a_m)$, write $\bar{a} = (a_2, \cdots, a_m)$. We shall now prove that, for $j - 1 \leqq ka_1 \leqq j$,

$$P_F\left\{S_n(j/k, \bar{a}) - F(j/k, \bar{a}) \geqq r\left(1 - \frac{2}{k-j}\right)n^{-1/2} \right. \tag{2.3}$$
$$\left. \mid X' = a;\ S_n(a) - F(a) \geqq rn^{-1/2};\ \bar{C}_{j-1,\,k}(r)\right\} \geqq 1 - \frac{(k-j)^2}{r} ,$$

which clearly implies (2.2).

Suppose the event conditioning (2.3) occurs. Since $\bar{C}_{j-1,k}(r)$ occurs, a_1 is actually $> (j-1)/k$, and there is w.p.1 at most one X_i with first coordinate a_1. Hence, if $\bar{l}$ denotes a vector consisting of $m-1$ ones, we have

$$nS_n(a_1, \bar{l}) \leqq nF(a_1, \bar{l}) + rn^{1/2} + 1 . \tag{2.4}$$

Hence, the number N of $X_1, \cdots, X_n$ which have first coordinate greater than a_1 is at least

$$M = n(1-a_1) - rn^{1/2} - 1 . \tag{2.5}$$

If $M \leqq 0$, we thus have $rn^{-1/2} \geqq -n^{-1} + (k-j)/k$, from which we obtain, for $r \geqq 1$,

$$\begin{aligned} &S_n(j/k, \bar{a}) - F(j/k, \bar{a}) \\ &\geqq [S_n(a_1, \bar{a}) - F(a_1, \bar{a})] + [F(a_1, \bar{a}) - F(j/k, \bar{a})] \\ &\geqq rn^{-1/2} - k^{-1} \geqq \left(1 - \frac{1}{k-j}\right) rn^{-1/2} - n^{-1}(k-j)^{-1} \\ &\geqq \left(1 - \frac{2}{k-j}\right) rn^{-1/2} . \end{aligned} \tag{2.6}$$

On the other hand, if $M > 0$ we have $N > 0$. Let Q be the event that, of the N random variables X_i whose first coordinate is $> a_1$, at least

$$n[F(j/k, \bar{a}) - F(a_1, \bar{a})] - 2n^{1/2}(k-j)^{-1}r$$

take on values in the region

$$\{(x_1, \cdots, x_m): a_1 < x_1 \leqq j/k;\ x_2 \leqq a_2, \cdots, x_m \leqq a_m\} .$$

Clearly, if we show that, for ν equal to any integer $\geqq M$,

$$\begin{aligned} P_F\{Q \mid X' = a;\ S_n(a) - F(a) \geqq rn^{-1/2};\ \bar{C}_{j-1,k}(r);\ N = \nu\} \\ \geqq 1 - \frac{(k-j)^2}{r^2} , \end{aligned} \tag{2.7}$$

this together with (2.6) yield (2.3).

Define

$$p = \frac{F(j/k, \bar{a}) - F(a_1, \bar{a})}{1 - a_1} .$$

If $p = 0$, (2.7) is trivial. We therefore assume $p > 0$, and define

$$t = \frac{n[F(j/k, \bar{a}) - F(a_1, \bar{a})] - 2n^{1/2}(k-j)^{-1}r - \nu p}{[\nu p(1-p)]^{1/2}} .$$

Since $\nu \geqq M$, (2.5) implies that

$$t \leqq \frac{r[p - 2(k-j)^{-1}] + pn^{-1/2}}{[\nu p(1-p)n^{-1}]^{1/2}} .$$

Since the event $X' = a$ depends only on $\{S_n(x), x_1 \leqq a_1\}$, the probability (2.7) is simply the probability that, of ν independent Bernoulli trials with probability p of a "success," the number Y of successes satisfies

$$\frac{Y - EY}{[E(Y - EY)^2]^{1/2}} \geqq t .$$

Since $p \leqq (k-j)^{-1}$ and $\nu/n \leqq 1$, it follows that $t \leqq -r(k-j)^{-1}$ for $r \geqq 2$. Hence, (2.7) follows at once from Chebyshev's inequality.

Next, let $B'_{jk}(r)$ and $C'_{jk}(r)$ be defined by replacing $S_n(x) - F(x)$ by $F(x) - S_n(x)$ on the right side of (2.1). In a manner almost identical to that used to prove Lemma 2, we obtain

LEMMA 2'. *We have, for* $r \geqq 2$,

$$P_F\left\{C'_{jk}\left(r\left(1 - \frac{2}{k-j}\right)\right) \mid \bar{C}'_{j-1,k}(r); B'_{jk}(r)\right\} \geqq 1 - \frac{(k-j)^2}{r^2} . \tag{2.8}$$

In fact, this is even easier, since the case $N = 0$ now requires no calculation. We replace (2.5) by an upper bound on N, Q by the event that at most $n[F(j/k, \bar{a}) - F(a_1, \bar{a})] + 2n^{1/2}r(k-j)^{-1}$ observations fall in the region of probability p, obtain a positive lower bound on t, and thus a lower bound on $P\{Y - EY \leqq t\,[E(Y - EY^2]^{1/2}\}$.

Our next lemma requires an induction on m, which we exhibit in its number.

LEMMA 3-m. *For* $\varepsilon > 0$ *there is a number* $c_1(\varepsilon, m)$ *such that, for all* F, n, $r \geqq 0$, $k \geqq 2$, *and* $j \leqq k$,

$$P_F\{C_{jk}(r) \bigcup C'_{jk}(r)\} < c_1(\varepsilon, m)e^{-(2-\varepsilon)r^2} . \tag{2.9}$$

Proof. Of course, for $m = 1$ the result (2.9) is weaker than (1.7) (or (1.1)). For $m > 1$, consider $S_n - F$ on the set $W_{jk} \bigcup \{x : x_1 > j/k, x_2 = 1\}$. This clearly has the same distribution theory as an $(m-1)$-dimensional sample d.f. minus the corresponding continuous d.f. The desired result thus follows from Theorem 1-$(m-1)$.

Since

$$P_F\{B_{jk}(\alpha)\} \leqq \frac{P_F\{C_{jk}(\beta)\}}{P_F\{C_{jk}(\beta) \mid \overline{C_{j-1,k}(\alpha)};\ B_{jk}(\alpha)\}} + P_F\{C_{j-1\,k}(\alpha)\}\ . \tag{2.10}$$

we obtain at once from Lemmas 2, 2′, and 3-m,

LEMMA 4-m. *For $\varepsilon' > 0$ there is a number $c_2(\varepsilon', m)$ such that, for all F, n, $k \geqq 2$, $j \leqq k-1$, and $r \geqq 2(k-j)$,*

$$P_F\left\{\sup_{V_{jk}} |S_n(x) - F(x)| \geqq rn^{-1/2}\right\} \leqq c_2(\varepsilon', m) \exp\left\{-r^2\left(1 - \frac{2}{k-j}\right)^2(2-\varepsilon')\right\}\ . \tag{2.11}$$

It now becomes evident that, by taking k large, we can prove Theorem 1-m by using the estimate (2.11) for those V_{jk} for which $k-j$ is large, provided we can also find an appropriate estimate for the region where $k-j$ is small, i.e., near $x_1 = 1$. Actually, we shall see that it suffices to find such an estimate for the region where x is close to $(1, 1, \cdots, 1)$, and the appropriate estimate is obtained in Lemma 6-m below. We first require a preliminary result which essentially improves the estimate (2.11) when j is small.

LEMMA 5-m. *There is a number $c_3(m)$ such that, for all F, $n \geqq 1$, $k \geqq 3$, and $r \geqq 2k[1 - 2/(k-1)]^{-m}$,*

$$P_F\left\{\sup_{V_{1k}} |S_n(x) - F(x)| \geqq rn^{-1/2}\right\} \leqq c_3(m)\exp\left\{-r^2\left[1 - \frac{4m}{k-1}\right]\log(ke^{-1})\right\}\ . \tag{2.12}$$

Proof. Let $T_m = V_{1k}$, $T_{m-1} = W_{1k}$, $T_{m-q} = \{x: x_1 = k^{-1}, x_2 = 1, \cdots, x_q = 1\}$ for $q > 1$. As in the proof of Lemma 3-m, we see that $S_n - F$ on T_{m-q} has the same distribution theory as an $(m-q)$-dimensional sample d.f. minus the corresponding continuous d.f. with uniform marginal d.f.'s, on the subset of the $(m-q)$-cube where the first coordinate is $\leqq k^{-1}$. Hence, applying the argument (2.10) m times, where the last term on the right side of (2.10) is now zero, where in each successive application we use Lemma 2 or 2′ to obtain 3/4 as a lower bound on the denominator of the expression on the right side, and where in the qth application the right side numerator and left side of (2.10) are, respectively, probabilities of maximum deviations greater than or equal to $r[1 - 2/(k-1)]^q n^{-1/2}$ on T_{m-q} and greater than or equal to $r[1 - 2/(k-1)]^{q-1}n^{-1/2}$ on T_{m-q+1}, we obtain

$$(2.13)\quad P_F\left\{\sup_{V_{1k}} |S_N(x) - F(x)| \geqq rn^{-1/2}\right\} \leqq \left(\frac{4}{3}\right)^m \phi(k^{-1}, n, r[1 - 2/(k-1)]^m),$$

where $\phi(p, n, z)$ is the probability that the number Z of successes in n independent Bernoulli trials with probability p of success satisfies

$$|Z - np| \geqq zn^{1/2}.$$

Lemma 1 and (2.13) thus imply (2.12).

We are now ready to prove that large deviations of $|S_n(x) - F(x)|$ have suitably small probability if x lies in the region

$$U_\delta = \{x : 1 - \delta \leqq x_q \leqq 1,\ 1 \leqq q \leqq m\}$$

and δ is a small positive number.

LEMMA 6-m. *There is a number $c_4(m)$ such that, for all F, n, $\delta \leqq 1/4$, and $r \geqq 2\delta^{-1}[1 - 3\delta]^{-m}$, we have*

$$(2.14)\quad P_F\left\{\sup_{U_\delta} |S_n(x) - F(x)| \geqq rn^{-1/2}\right\} \leqq c_4(m)\exp\{-r^2 4^{-m}(1 - 6m\delta)\log(\delta^{-1}e^{-1})\}.$$

Proof. For a in U_δ, let $Q_a = \{x: x_q \leqq a_q,\ 1 \leqq q \leqq m\}$, and for any sequence $\sigma = (\sigma_1, \cdots, \sigma_m)$ of 1's and -1's, not all 1's, let

$$Q_a(\sigma) = \{x: x \in I^m;\ \sigma_q x_q < \sigma_q a_q,\ 1 \leqq q \leqq m\}.$$

Thus, Q_a and the $2^m - 1$ sets $Q_a(\sigma)$ are disjoint sets whose union is I^m minus the union of m hyperplanes $\{x: x_q = a_q\}$. Define the event

$$D_a(\sigma, s) = \left\{|[\text{number of } X_1, \cdots, X_n \text{ falling in } Q_a(\sigma)] - n\int_{Q_a(\sigma)} dF| \geqq n^{1/2}s\right\}.$$

If the event

$$B_\delta(r) = \left\{\sup_{U_\delta} |S_n(x) - F(x)| \geqq rn^{-1/2}\right\}$$

occurs, we define, in the manner of the definition of the random vector X' of the proof of Lemma 2, a random vector X'' such that $X'' \in U_\delta$ and $|S_n(X'') - F(X'')| \geqq rn^{-1/2}$. With probability one there is at most one X_i with any coordinate equal to the corresponding coordinate of X'', so that w.p. 1 the event $B_\delta(r) \cap \{X'' = a\}$ entails the event

$$(2.15)\quad \bigcup_\sigma D_a(\sigma, r/(2^m - 1)).$$

For any fixed σ, let $Y_i = (Y_i^{(1)}, \cdots, Y_i^{(m)})$ be defined by $Y_i^{(q)} = X_i^{(q)}$ (resp., $= 1 - X_i^{(q)}$) if $\sigma_q = 1$ (resp., $= -1$). Let G be the d.f. of Y_1 when F is the d.f. of X_1, and let S_n' be the sample d.f. of $Y_1, \cdots, Y_n$.

Let $b_q(\sigma, \delta) = 1 - \delta$ (rep., δ) if $\sigma_q = 1$ (resp., $= -1$), and let $V(\sigma, \delta) = \{x: 0 \leqq x_q \leqq b_q(\sigma, \delta),\ 1 \leqq q \leqq m\}$. The event $\bigcup_{a \in U_\delta} D_a(\sigma, s)$ is, w.p. 1, equivalent to the event

$$\left\{\sup_{V(\sigma, \delta)} |S'_n(x) - G(x)| \geqq s\right\}. \tag{2.16}$$

Since at least one σ_q is -1, at least one $b_q(\sigma, \delta)$ equals δ. Hence, if $k \geqq \delta^{-1}$, the set $V(\sigma, \delta)$ is a subset of a set obtained from V_{1k} by relabeling coordinates. Hence, by Lemma 5-m, the event (2.16) has probability no greater than

$$c_3(m)\exp\{-s^2[1 - 6m\delta]\log(\delta^{-1}e^{-1})\}.$$

The union (over σ) of $2^m - 1$ such events (with $s = r/(2^m - 1)$) is equivalent to the union over a in U_δ of the events (2.15), and thus contains the event $B_\delta(r)$, w.p. 1. This proves (2.14).

Proof of Theorem 1-m. (The proof which follows is valid for $m \geqq 1$, and does not require the use of (1.1), although the latter implies the desired result when $m = 1$.) Suppose $m \geqq 1$, and let $\varepsilon > 0$ be given. Choose δ to be the largest positive number which is $\leqq 1/6m$ and such that $4^{-m}(1 - 6m\delta)\log(\delta^{-1}e^{-1}) > 2$. Let k be the smallest positive number such that $k^{-1} < \delta/2$ and such that $2/(k - j) < \varepsilon/5$ for $j/k < 1 - \delta/2$, and let $\varepsilon' = \varepsilon/5$ in (2.11). The coefficients of $-r^2$ in the exponentials of (2.14) and (2.11) are thus $\geqq 2 - \varepsilon$ provided $j/k < 1 - \delta/2$. Thus, writing $V_q = \{x: x_q \leqq 1 - \delta\}$, we see that V_1 is contained in a union of fewer than k sets V_{jk} (namely, those for which $j/k < 1 - \delta/2$), for each of which the coefficient of $-r^2$ in (2.11) is $\geqq 2 - \varepsilon$. Thus, $r \geqq R(\varepsilon, m)$ implies

$$P_F\left\{\sup_{V_1} |S_n(x) - F(x)| \geqq rn^{-1/2}\right\} \leqq c_5(\varepsilon, m)e^{-r^2(2-\varepsilon)}, \tag{2.17}$$

where R and c_5 depend on ε and m but not on r, F, or n. Interchanging the roles of the first and qth coordinates, we obtain (2.17) with V_1 replaced by V_q. Since the union of U_δ and the V_q's ($1 \leqq q \leqq m$) is I^m, we obtain (1.5) for $r \geqq R'(\varepsilon, m)$ and thus (by possibly increasing $c(\varepsilon, m)$) for all $r \geqq 0$.

REFERENCES

1. K. L. Chung, *An estimate concerning the Kolmogoroff limit distribution*, Trans. Amer. Math. Soc., **67** (1949), 36-50.

2. A. Dvoretzky, J. Kiefer, and J. Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, Ann. Math. Stat., **27** (1956), 642-669.

3. W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, 2nd

edition, Wiley, New York (1957).
4. J. Kiefer and J. Wolfowitz, *On the deviations of the empiric distribution function of vector chance variables*, Trans. Amer. Math. Soc., **87** (1958), 173-186.
5. A. N. Kolmogorov, *Sulla determinrzione empirica di una legge di distribuzione*, Ist. Ital. Atti. Giorn., **4** (1933), 83-91.
6. P. Lévy, *Théorie de L'Addition des Variales Aléatoires*, Gauthier-Villard, Paris (1937).

CORNELL UNIVERSITY

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is $12.00; single issues, $3.50. Back numbers are available. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $4.00 per volume; single issues, $1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 103 Highland Boulevard, Berkeley 8, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

The Supporting Institutions listed above contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its content or policies.

Pacific Journal of Mathematics
Vol. 11, No. 2 December, 1961