# THE PROBABILITY IN THE EXTREME TAIL OF A CONVOLUTION[1]

DAVID BLACKWELL AND J. L. HODGES, JR.

*University of California, Berkeley*

**1. Summary.** Let $X_1, X_2, \cdots$ be independent and identically distributed random variables with possible values that are integers whose differences have g.c.d. one. Assume the m.g.f. of $X_1$ exists in an interval about 0, let $a$ be any number such that $E(X_1) < a < \sup X_1$, and let $\phi(a, t) = Ee^{t(X_1-a)}$. There exists a unique value $t^*(a)$ of $t$ which minimizes $\phi(a, t)$ with respect to $t$; write $m(a) = \phi[a, t^*(a)]$ and $z = e^{-t^*(a)}$. Let $Y_1, Y_2, \cdots$ be independent and identically distributed random variables such that $Y_1$ and $X_1$ have the same range and $\Pr(Y_1 = x) = \Pr(X_1 = x)\cdot e^{t^*(a)(x-a)}/m(a)$, and let $\mu_2 = \sigma^2, \mu_3, \mu_4$ be central moments of $Y_1$.

We show that $\Pr\{X_1 + \cdots + X_n = na\} = [m(a)]^n \Pr\{Y_1 + \cdots + Y_n = na\}$, and use this to establish the approximation $\Pr\{X_1 + \cdots + X_n = na\} = \pi_n^{**}[1 + 0(n^{-2})]$, where $na$ is a possible value of $X_1 + \cdots + X_n$ and

$$\pi_n^{**} = \frac{[m(a)]^n}{\sigma\sqrt{2\pi n}}\left[1 + \frac{1}{8n}\left(\frac{\mu_4}{\mu_2^2} - 3 - \frac{5}{3}\frac{\mu_3^2}{\mu_2^3}\right)\right].$$

Similarly we find that $\Pr\{X_1 + \cdots + X_n \geqq na\} = \Pi_n^{**}[1 + 0(n^{-2})]$, where

$$\Pi_n^{**} = \pi_n^{**}\cdot\frac{1}{1-z}\left\{1 - \frac{1}{2n}\left[\frac{(z\mu_3/\mu_2) + z(1+z)/(1-z)}{(1-z)\mu_2}\right]\right\}.$$

We provide some numerical illustrations of the accuracy of these approximations, and give a conjectured analog of the leading term of $\Pi_n^{**}$ for nonlattice variables.

**2. Introduction.** Let $X_1, X_2, \cdots$ be independent identically distributed random variables whose common moment generating function $Ee^{tX_1}$ is finite in some interval about 0, and let $a$ be any number such that $E(X_1) < a < \sup X_1$. We shall be interested in the tail probability

$$\Pi_n(a) = \Pr\{X_1 + \cdots + X_n \geqq na\}.$$

As $n \to \infty$ we shall of course have $\Pi_n(a) \to 0$, since $na$ exceeds the expected value of the sum by about $\sqrt{n}$ standard deviations. The study of the speed with which $\Pi_n(a) \to 0$ was initiated by Cramér [2] in 1938; his results were extended by Feller [3] and Chernoff [1].

Denote by $\phi(a, t)$ the moment generating function of $X_1 - a$: $\phi(a, t) = Ee^{t(X_1-a)}$. Chernoff shows that for each $a$ there is a unique value of $t$, say $t^*(a)$,

Received September 26, 1958; revised August 3, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53) and (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.



Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

for which $\phi$ achieves its minimum, and writes $\phi[a, t^*(a)] = m(a)$. He shows that. for every $\epsilon > 0$,

$$[m(a) - \epsilon]^n \leqq \Pi_n(a) \leqq [m(a)]^n$$

for sufficiently large $n$, with the right inequality holding for all $n$.

This result establishes in a sense the speed with which $\Pi_n(a) \to 0$, but it is not precise enough to permit the approximation of $\Pi_n(a)$ with a small relative error, since the ratio of upper to lower bound tends to infinity with $n$. There remains the problem of developing a relatively accurate approximation for $\Pi_n(a)$. Cramér [2] has found such an approximation for the case in which $X_1$ has an absolutely continuous component. We are interested in the case of lattice variables, i.e., the case in which there are constants $A \neq 0$ and $B$ such that $AX_1 + B$ has only integer values.

**3. An identity.** In this section we restrict attention to sequences $\{X_n\}$ of discrete variables.

THEOREM 1: *Let $X_1, X_2, \cdots$ be independent identically distributed discrete variables whose common moment generating function $E(e^{tX_1})$ is finite for some interval about* 0. *For any $a$ with $E(X_1) < a < \sup X_1$, let*

$$m(a) = \min_t Ee^{t(X_1 - a)} = \min_t \phi(t, a) = \phi[t^*(a), a], \text{ say,}$$

*and let $Y_1, Y_2, \cdots$ be independent identically distributed discrete variables whose common distribution is defined by*

$$\Pr\{Y_1 = x\} = \Pr\{X_1 = x\} \exp [t^*(a)(x - a)]/m(a) \quad \text{for all } x.$$

*Then for all $n$,*

$$\Pr\{X_1 + \cdots + X_n = na\} = [m(a)]^n \Pr\{Y_1 + \cdots + Y_n = na\}.$$

The shift from the random variable $X$ to the random variable $Y$ has the effect of moving our event from the extreme tail to the center, since $na$ is just the expected value of $Y_1 + \cdots + Y_n$. This shift is not new. It is essentially carried out in Cramér's original paper. Wald [6] made a similar change in his "conjugate" distribution, introduced in the study of a problem arising in sequential analysis. Shannon [4] encountered the shift in a problem of information theory, and remarked (p. 15): "These tilted probabilities are convenient in evaluating the 'tails' of distribution that are sums of other distributions."

PROOF OF THEOREM 1: As noted by Chernoff, $\phi(t, a)$ is for each $a$ a strictly convex function of $t$ and attains its minimum at a unique $t = t^*(a)$. Write $p(x) = \Pr\{X_1 = x\}$. We have $\phi(a, t) = \sum_x p(x)e^{t(x-a)}$, so that

$$\phi_2[a, t^*(a)] = \sum_x (x - a)p(x)e^{t^*(a)(x-a)} = 0, \tag{1}$$

where $\phi_i$ denotes the partial derivative of $\phi$ with respect to its $i$th argument. Write $q(x) = p(x)e^{t^*(a)(x-a)}/m(a)$. Then $q(x)$ is a discrete probability distribution, and (1) asserts that the mean of the $q$ distribution is $a$. Let $Y_1, Y_2, \cdots$

be a sequence of independent identically distributed variables with common distribution $q$, and let $x_1, \cdots, x_n$ be any sequence of numbers whose sum is $na$. Then

$$\begin{aligned}\Pr\{(Y_1, \cdots, Y_n) = (x_1, \cdots, x_n)\} &= q(x_1) \cdots q(x_n) \\ &= p(x_1) \cdots p(x_n) \exp [t^*(a)(x_1 + \cdots + x_n - na)]/[m(a)]^n \\ &= \Pr\{(X_1, \cdots, X_n) = (x_1, \cdots, x_n)\}/[m(a)]^n.\end{aligned}$$

Summing over all sequences $(x_1, \cdots, x_n)$ such that $x_1 + \cdots + x_n = na$ yields the assertion of the theorem.

Theorem 1 extends to $M$-dimensional variables without change. We shall not use this extension, but give it for completeness.

THEOREM 2: *Let $X_1, X_2, \cdots$ be independent identically distributed discrete $M$-dimensional variables, and let $a$ be any interior point of the convex hull of the range of $X_1$, $a \neq \mu$, where $\mu = E(X_1)$. Suppose that there is a positive number $b$ such that the moment generating function $Ee^{t\cdot X_1}$ is finite for all $t$ for which $|t| \leqq b$ and $t\cdot(a - \mu) \geqq 0$ where, for any $t = (t_1, \cdots, t_m)$, $x = (x_1, \cdots, x_m)$, $t\cdot x$ denotes the inner product $\sum t_i x_i$. Then the moment generating function of $X_1 - a$ achieves its minimum value $m(a)$, say, at a unique $t = t^*(a)$, say, and, if $Y_1, Y_2, \cdots$ are independent identically distributed discrete $M$-dimensional variables whose common distribution is defined by* $\Pr\{Y_1 = x\} = \Pr\{X_1 = x\}e^{t^*\cdot(x-a)}/m(a)$, *then, $Y_1$ has mean $a$ and, for all $n$,*

$$\Pr\{X_r + \cdots + X_n = na\} = [m(a)]^n \Pr\{Y_1 + \cdots + Y_n = na\}.$$

The proof parallels that of Theorem 1. Again, the moment generating function has a minimum $m(a)$ at a unique $t^*$, at which $\partial\phi \mid \partial t_i = 0$ for all $i$. These equations assert that the $q$ distribution defined by $q(x) = \Pr(X_1 = x)e^{t^*\cdot(x-a)}/m(a)$ has mean $a$, and the rest of the proof is as before.

**4. The individual term.** In this section we shall specialize to the case of lattice variables. This means that it is possible by a linear transformation to assure that the values of $X_1$ are integers whose differences have g.c.d. 1; we assume this reduction has been carried out. We are then able to develop expressions for $\pi_n(a)$, using a method exploited for example by von Mises [5 Sec. 8].

Let $\sigma^2 = \mu_2$, $\mu_3$, $\mu_4$ be central moments of $Y$ of order 2, 3, 4. We shall establish

THEOREM 3: *If $X_1, X_2, \cdots$ are integer-valued variables satisfying the hypotheses of Theorem 1, the approximation*

$$\pi_n^*(a) = \frac{[m(a)]^n}{\sqrt{2\pi n}\sigma}$$

*for $\pi_n(a) = \Pr\{X_1 + \cdots + X_n = na\}$ has relative error of order $n^{-1}$, while the approximation*

$$\pi_n^{**}(a) = \pi_n^*(a)\left\{1 + \frac{1}{8n}\left[\frac{\mu_4}{\mu_2^2} - 3 - \frac{5}{3}\frac{\mu_2^3}{\mu_2^3}\right]\right\}$$

*for $\pi_n(a)$ has relative error of order $n^{-2}$.*

PROOF: In general, if a random variable $U$ with characteristic function $\eta$ has only integral values it is easy to check [5] that

$$\Pr(U = u) = (1/2\pi) \int_{-\pi}^{\pi} e^{-itu}\eta(t)\, dt.$$

Since $Y_1 + \cdots + Y_n$ is such a random variable, we have

$$\Pr(Y_1 + \cdots + Y_n = na) = (1/2\pi) \int_{-\pi}^{\pi} e^{-itna}\zeta^n(t)\, dt$$

where $\zeta(t)$ is the characteristic function of $Y$ and $na$ is an integer. Finally, if we write $\psi(t) = e^{-iat}\zeta(t)$ for the characteristic function of $Y - a$, we have $\Pr(Y_1 + \cdots + Y_n = na) = (1/2\pi) \int_{-\pi}^{\pi} \psi^n(t)\, dt$.

To evaluate this integral, let us first take it over the range $|t| \leqq \log n/\sqrt{n}$. If we make the usual expansion of $\log \psi(t)$ in terms of the cumulants $\kappa_r$ of $Y_1 - a$, observe $\kappa_1 = 0$, and write $\kappa_2 = \sigma^2$, we find

$$\psi^n(t) = e^{-\frac{n\sigma^2 t^2}{2}} \exp\left\{ n \sum_{r=3}^{6} \frac{\kappa_r (it)^r}{r!} + o(n^{-2}) \right\}$$

when $|t| \leqq \log n/\sqrt{n}$. The transformation $\sqrt{n}\,\sigma\, t = u$ and series expansion of the second factor puts the integrand in the form

$$(2) \qquad e^{-u^2/2}\left\{ 1 - \frac{i\kappa_3 u^3}{6\sigma^3\sqrt{n}} + \frac{1}{n}\left[\frac{\kappa_4 u^4}{24\sigma^4} - \frac{\kappa_3^2 u^6}{72\sigma^6}\right] + \frac{uP_1}{n^{3/2}} + \frac{P_2}{n_2} + o(n^{-2})\right\}$$

over $|u| \leqq \sigma \log n$, where $P_\iota$ denotes a polynomial in $u^2$. Using the fact that

$$(3) \qquad \int_{-\sigma \log n}^{\sigma \log n} u^p e^{-u^2/2}\, du = 2^{\frac{p+1}{2}}\, \Gamma\left(\frac{p+1}{2}\right) + o(n^{-2})$$

when $p$ is even, and vanishes when $p$ is odd, we find

$$(4) \qquad \frac{1}{2\pi} \int_{-\log n/\sqrt{n}}^{\log n/\sqrt{n}} \psi^n(t)\, dt = \frac{1}{\sigma\sqrt{2\pi n}} \left\{1 + \frac{1}{8n}\left[\frac{\mu_4}{\mu_2^2} - 3 - \frac{5\mu_3^2}{3\mu_2^3}\right] + o(n^{-2})\right\}$$

where we have expressed the cumulants in terms of the central moments $\mu_r$.

Turning now to the range $\log n/\sqrt{n} \leqq |t| \leqq \pi$, we shall show that this part of the integral is negligible. Since $\kappa_1 = 0$ and $0 < \sigma^2 < \infty$, we can find $0 < t_0 < \pi$ such that $|\psi(t)| \leqq 1 - (\sigma^2 t^2/3)$ for $|t| \leqq t_0$. Therefore, over the range $\log n/\sqrt{n} \leqq |t| \leqq t_0$,

$$\left| \int \psi^n(t) dt \right| \leqq 2 \int_{\log n/\sqrt{n}}^{\infty} e^{-n\sigma^2 t^2/3}\, dt,$$

which is $o(n^{-k})$ for all $k$. As for $t_0 \leqq |t| \leqq \pi$, note first that our assumption that the possible values of $X_1$ are integers whose differences have g.c.d. 1 implies that, when $0 < |t| \leqq \pi$, the points $e^{itx}$ can never all coincide, and hence that $\sum_x q(x)e^{itx}$ lies inside the unit circle. Therefore

$$|\psi(t)| = |e^{-ita} \sum_x q(x)e^{itx}| < 1 \quad \text{for} \quad t_0 \leqq |t| \leqq \pi,$$

and by the continuity of $\psi$ there is a number $\rho < 1$ for which $|\psi(t)| < \rho$ in this range, over which $\int \psi^n(t)\,dt$ is $o(\rho^n)$. We may therefore take the right side of (4) as an expression for $(1/2\pi) \int_{-\pi}^{\pi} \psi^n(t)\,dt$, and hence for

$$\Pr(Y_1 + \cdots + Y_n = na).$$

This fact, combined with Theorem 1, proves Theorem 3.

We present in Table 1 a few illustrations of the accuracy of the two approximations. Here, by the relative error of an approximation $\pi'$ for a quantity $\pi$ we mean $(\pi'/\pi) - 1$. The values of $X$ are $0, 1, \cdots, r - 1$.

**5. The tail probability.** An extension of the methods used above provides expressions for the tail probability. We are indebted to D. A. Darling for suggestions which led to this result.

THEOREM 4: *If $X_1, X_2, \cdots$ are integer-valued random variables satisfying the hypotheses of Theorem 1, then the approximation*

$$\Pi_n^*(a) = \pi_n^*(a)/(1 - z)$$

*for $\Pi_n(a) = \Pr\{X_1 + \cdots + X_n \geqq na\}$ has relative error of order $n^{-1}$, while the approximation*

$$\Pi_n^{**}(a) = \Pi_n^*(a)\left\{1 - \frac{1}{2n}\left[\frac{(z\mu_3/\mu_2) + z(1 + z)/(1 - z)}{(1 - z)\mu_2}\right]\right\}$$

*for $\Pi_n(a)$ has relative error of order $n^{-2}$.*

PROOF: An easy modification of Theorem 1 shows that, for any integer $k$,

$$\begin{aligned}\pi(k) &= \Pr(X_1 + \cdots + X_n = na + k)\\ &= [m(a)]^n e^{-kt^*} \Pr(Y_1 + \cdots + Y_n = na + k)\end{aligned}$$

TABLE 1

| $r$ | $p$ | $a$ | $n$ | $\pi_n$ | Relative error of | |
|---|---|---|---|---|---|---|
| | | | | | $\pi_n^*$ | $\pi_n^{**}$ |
| 3 | $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ | $\frac{3}{2}$ | 8 | .040542600 | .0327 | $-.0^3173$ |
| | | | 16 | $.0^260067293$ | .0162 | $.0^4252$ |
| | | | 32 | $.0^318201692$ | $0.0^2804$ | $.0^5740$ |
| | | | 64 | $.0^623350658$ | $0.0^2400$ | $.0^5203$ |
| | | $\frac{7}{4}$ | 8 | $.0^254869684$ | .0556 | $-.0^3816$ |
| | | | 16 | $.0^484094675$ | .0275 | $.0^4654$ |
| | | | 32 | $.0^726832893$ | .0101 | $.0^4192$ |
| | $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ | $\frac{3}{2}$ | 8 | $.0^279040527$ | .0331 | $-.0^2379$ |
| | | | 16 | $.0^324029752$ | .0182 | $-.0^4188$ |
| | | | 32 | $.0^630784454$ | $.0^2901$ | $.0^5307$ |
| 4 | $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ | $\frac{9}{4}$ | 8 | .021789551 | .0291 | $.0^3395$ |
| | | | 16 | $.0^223290971$ | .0142 | $.0^3109$ |

while the proof of Theorem 3 gives

$$\Pr(Y_1 + \cdots + Y_n = na + k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \psi^n(t)\, dt.$$

Summation over $k$ now gives

$$\Pi_n = \lim_{K\to\infty} \sum_{k=0}^{K} \pi(k) = \frac{[m(a)]^n}{2\pi} \lim_{K\to\infty} \int_{-\pi}^{\pi} \frac{1 - e^{-K(it+t^*)}}{1 - e^{-(it+t^*)}} \psi^n(t)\, dt.$$

Because of the boundedness of the integrand, we may pass to the limit inside the integral to get

$$\Pi_n = \frac{[m(a)]^n}{2\pi} \int_{-\pi}^{\pi} \frac{1}{1 - ze^{-it}} \psi^n(t)\, dt. \tag{5}$$

where $z = e^{-t^*} < 1$.

The evaluation of this integral is much like that in the proof of Theorem 3. Since $1/(1 - ze^{-it})$ is bounded, the integral over $|t| \geqq \log n/\sqrt{n}$ is negligible as before. As before, we substitute $\sqrt{n}\,\sigma t = u$, and find that when $|u| \leqq \sigma \log n$,

$$\frac{1}{1 - ze^{-it}} = \frac{1}{1-z} - \frac{izu}{\sigma(1-z)^2\sqrt{n}} - \frac{z(1+z)u^2}{2\sigma^2(1-z)^3 n} + \frac{uP_3}{n^{3/2}} + \frac{P_4}{n^2} + o(n^{-2})$$

where the $P_i$ again denote polynomials in $u^2$. Combining this with (2), and integrating the various terms with the aid of (3), we find

$$\int_{-\pi}^{\pi} \frac{1}{1 - ze^{-it}} \psi^n(t)\, dt = \frac{\sqrt{2\pi}}{1-z}\left\{1 + \frac{1}{8n}\left[\frac{\mu_4}{\mu_2^2} - 3 - \frac{5\mu_3^2}{3\mu_2^3}\right] - \frac{z}{2n}\frac{\mu_3(1-z) + \sigma^2(1+z)}{\sigma^4(1-z)^2} + o(n^{-2})\right\}.$$

This, combined with (5), yields Theorem 4.

We present in Table 2 a few illustrations of the accuracy of the approximations. As in Table 1, the values of $X$ are 0, 1, $\cdots$, $r - 1$.

TABLE 2

| $r$ | $p$ | $a$ | $n$ | $\Pi_n$ | Relative error of $\Pi_n^*$ | Relative error of $\Pi_n^{**}$ |
|---|---|---|---|---|---|---|
| 3 | $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ | $\frac{3}{2}$ | 8 | .064471879 | 0.148 | −.0888 |
| | | | 16 | $.0^299484233$ | 0.0846 | −.0289 |
| | | | 32 | $.0^330990382$ | 0.0465 | −.00861 |
| 3 | $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$ | $\frac{3}{2}$ | 8 | .011276245 | 0.0862 | −.0259 |
| | | | 16 | $.0^335039405$ | 0.0474 | $-.0^2733$ |
| 4 | $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ | $\frac{9}{4}$ | 8 | .040328979 | 0.134 | −.0705 |
| | | | 16 | $.0^244785112$ | 0.0757 | −.0224 |

**6. The nonlattice case.** For nonlattice variables, let us heuristically treat $\pi_n^*(a) = m^n(a)/\sigma(a)\sqrt{2\pi n}$ as an approximation to the (in general nonexistent) density of $X_1 + \cdots + X_n$ at the point $na$, and proceed formally.

$$\begin{aligned}
\Pr\{X_1 + \cdots + X_n \geqq na\} &\sim \int_0^\infty \pi_n^*\left(a + \frac{x}{n}\right) dx \\
&= \pi_n^*(a) \int_0^\infty \left[\pi_n^*\left(a + \frac{x}{n}\right) \Big/ \pi_n^*(a)\right] dx \\
&\sim \pi_n^*(a) \int_0^\infty \left[\frac{m\left(a + \frac{x}{n}\right)}{m(a)}\right]^n dx \\
&\sim \pi_n^*(a) \int_0^\infty \exp\left[xm'(a)/m(a)\right] dx \\
&= \pi_n^*(a) \int_0^\infty \exp\left[-xt^*(a)\right] dx = \pi_n^*(a)/t^*(a).
\end{aligned}$$

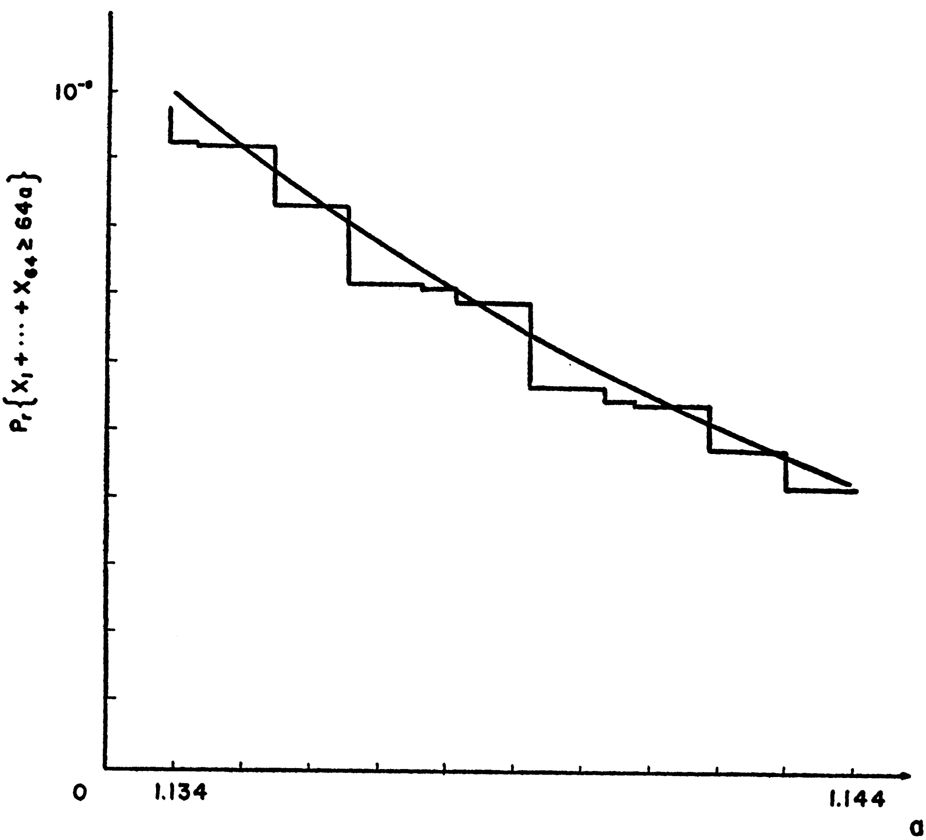


FIG. 1

We thus obtain the approximation

$$\Pi_n^*(a) = \pi_n^*(a)/t^*(a).$$

We conjecture that this approximation has a relative error which is $O(n^{-1})$, just as the corresponding approximation did in the lattice case. For variables with an absolutely continuous component, $\Pi_n^*(a)$ is just the leading term in the expansion obtained by Cramér [2] and is thus known to be correct. The conjecture is supported by numerical evidence for the case in which $X$ has values 0, 1, and $\sqrt{2}$ with equal probabilities. We have computed a portion of the tail of this distribution for $n = 64$, which is shown in Fig. 1 with the approximation superimposed.

## REFERENCES

[1] HERMAN CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[2] H. CRAMÉR, "Sur un noveau théorème-limite de la théorie des probabilités," *Actualités Scientifiques et Industrielles*, No. 736, Paris, 1938.

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

[4] C. E. SHANNON, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1957), pp. 6–25.

[5] RICHARD VON MISES, "Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik," Leipzig 1931.

[6] ABRAHAM WALD, "Some generalizations of the theory of cumulative sums of random variables," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 287–293.