Adv. Appl. Prob. **12**, 475–490 (1980)
Printed in N. Ireland
0001-8678/80/020475-16$1.85

SADDLE POINT APPROXIMATION FOR THE DISTRIBUTION OF THE SUM OF INDEPENDENT RANDOM VARIABLES

ROBERT LUGANNANI* AND
STEPHEN RICE,* *University of California, San Diego*

Abstract

In the present paper a uniform asymptotic series is derived for the probability distribution of the sum of a large number of independent random variables. In contrast to the usual Edgeworth-type series, the uniform series gives good accuracy throughout its entire domain. Our derivation uses the fact that the major components of the distribution are determined by a saddle point and a singularity at the origin. The analogous series for the probability density, due to Daniels, depends only on the saddle point. Two illustrative examples are presented that show excellent agreement with the exact distributions.

SADDLE POINT APPROXIMATION; SUM OF INDEPENDENT RANDOM VARIABLES; UNIFORM ASYMPTOTIC SERIES

1. Introduction

The problem of calculating the probability $Q_N(y)$ that the sum

$$Y = v_1 + v_2 + \cdots + v_N \tag{1}$$

of N independent, identically distributed, random variables will exceed y has been extensively studied. In technical applications, where numerical values are of prime importance, a number of methods of determining $Q_N(y)$ have been used. A common one is to use the fast Fourier transform which works well when $Q_N(y)$ is not too small or when N is not too large. Another is to evaluate numerically the integral

$$Q_N(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iuy} [g(u)]^N \, du/(iu),$$

where the characteristic function $g(u)$ is the Fourier transform of the probability density $p_1(v)$ of the typical v_j in (1), and the path of integration is indented

Received 9 February 1978; revision received 20 April 1979.
* Postal address: Department of Applied Physics and Information Science, University of California, San Diego, La Jolla, CA 92093, U.S.A.
This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR 74–2689.

downwards at the origin. This method is capable of high accuracy but the integral often converges slowly and a detailed study of the asymptotic behavior of $g(u)$ may be required.

Here we present an asymptotic series for $Q_N(y)$ that we have found useful in calculations associated with signal detection problems (see the remarks made by Olver on the philosophy of using asymptotic series for numerical calculations [6], p. 519). Our series takes into account, in the manner of uniform asymptotic series, the mutual effect of the pole of the integrand at $u=0$ and the 'principal saddle point' u_0 on the imaginary u-axis. Although u_0 does not exist for all densities $p_1(v)$, it does exist in many cases of practical interest. The question of existence has been studied by Daniels [5].

The saddle point u_0 has been used in a number of investigations. It appears, in effect, in the study of large deviations. See Petrov [7], Chapter 8 where work by Cramér, Saulis and others is described. Daniels [5] has given an asymptotic series for the probability density $p_N(y)$ of (1) based on u_0, whose integration provides an approach alternative to ours (cf. Part (d) of Section 3). Roberts [9] has used u_0 to deal with communication problems and it appears implicitly in the Chernoff bound [4].

The series for $Q_N(y)$ is described in Section 2. In Section 3 several remarks are made about the series and the existence of u_0 is discussed briefly. Section 4 gives sufficient conditions for our series to be truly asymptotic. Estimation of the error is discussed in Section 5 and illustrated by examples in Sections 6 and 7. In Section 6 an example in which $p_1(v)$ is an exponential density is discussed and in Section 7 the uniform density is examined. Finally, in Section 8 the results are used to compare values of $Q_N(y)$ obtained from our series with those obtained from formulas given by Cramér and Saulis.

2. The asymptotic series for $Q_N(y)$

The integral for $Q_N(y)$ can be written as

$$Q_N(y)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{N[\phi(iu)-iur]}\,du/(iu), \tag{2}$$

where $\exp[\phi(iu)]=g(u)$ and $r=y/N$. Our main result is the asymptotic series

$$Q_N(y)\sim\tfrac{1}{2}\operatorname{erfc}(\sqrt{-f_0})+\sum_{n=0}^{\infty}(A_n-B_n), \tag{3}$$

$$f_0=N\phi(iu_0)-iu_0y, \tag{4}$$

$$\tfrac{1}{2}\operatorname{erfc}(x)=\pi^{-\frac{1}{2}}\int_x^{\infty}\exp(-t^2)\,dt=(2\pi)^{-\frac{1}{2}}\int_{x\sqrt{2}}^{\infty}\exp(-t^2/2)\,dt,$$

where A_n is given below by (9) for $n=0$, 1, 2, and B_n is given by (8). The series (3) is derived in the appendix. In (4) u_0 is the principal saddle point of $\exp[N\phi(iu)-iuy]$ mentioned in the introduction. If this saddle point exists it lies on the imaginary u-axis and is the root of

$$\frac{d\phi(iu)}{du}-ir=0 \tag{5}$$

which becomes zero when $y=\bar{y}=EY$.

Equation (3) is a special case of a class of 'uniform asymptotic series' for integrals containing a large parameter (Bleistein [2], van der Waerden [11], and Rice [8]). The large parameter is N and the uniformity is with respect to r.

In calculating A_n and B_n it is convenient to set $t_0=iu_0$ where t_0 is the appropriate real root of

$$\frac{d}{dt}\phi(t)-r=0. \tag{6}$$

It turns out that t_0 is positive when $y>\bar{y}$, negative when $y<\bar{y}$, and zero when $y=\bar{y}$. In terms of t_0 (4) becomes

$$f_0=N\phi(t_0)-t_0y, \tag{7}$$

where it can be shown that $f_0 \leqq 0$ with equality only when $t_0=0$. The sign of $\sqrt{-f_0}$ is taken to be the same as that of t_0.

The term B_n is the nth term in the asymptotic series for $\frac{1}{2}$ erfc $(\sqrt{-f_0})$,

$$B_n=\tfrac{1}{2}(-\pi f_0)^{-\frac{1}{2}}f_0^{-n}(\tfrac{1}{2})_n \exp(f_0), \tag{8}$$

where $(\alpha)_0=1$ and $(\alpha)_n=\alpha(\alpha+1)\cdots(\alpha+n-1)$ when $n>0$. The term A_n is the nth term in the asymptotic series obtained using the classical saddle point method to expand the integral (2) for $Q_N(y)$ about u_0. As $(y-\bar{y})/N$ increases, the distance between u_0 and the pole at $u=0$ increases, and the classical asymptotic series $Q_N(y)\sim\sum A_n$ becomes increasingly accurate. The presence of the complementary error function and the terms B_n increase the accuracy of (3) for small values of $(y-\bar{y})/N$.

The first three values of A_n are

$$\begin{aligned}
A_0 &= \mu[2\pi N]^{-\frac{1}{2}}\exp(f_0),\\
A_1 &= -3A_0N^{-1}[\tfrac{1}{3}\mu^2+\mu\theta_3+\tfrac{1}{2}(5\theta_3^2-2\theta_4)],\\
A_2 &= 15A_0N^{-2}[\tfrac{1}{5}\mu^4+\mu^3\theta_3+\tfrac{1}{2}\mu^2(7\theta_3^2-2\theta_4)\\
&\quad+\tfrac{1}{4}\mu(42\theta_3^3-28\theta_3\theta_4+4\theta_5)\\
&\quad+\tfrac{1}{8}(231\theta_3^4-252\theta_3^2\theta_4+56\theta_3\theta_5+28\theta_4^2-8\theta_6)],
\end{aligned} \tag{9}$$

where

$$\begin{aligned}\phi^{(n)} &= [(d/dt)^n \phi(t)]_{t_0} = [i^{-n}(d/du)^n \phi(iu)]_{u_0},\\ \theta_n &= \phi^{(n)}/(n![\phi^{(2)}]^{n/2}),\\ \mu &= 1/(t_0[\phi^{(2)}]^{\frac{1}{2}}).\end{aligned} \tag{10}$$

Here $\phi^{(2)}$ is positive and $\phi^{(n)}$ can be interpreted as the nth cumulant of the 'associated' density (see Daniels [5], pages 639 and 640)

$$\begin{aligned}\hat{p}_1(v) &= e^{vt_0} p_1(v)/g(u_0),\\ g(u_0) &= \int_{-\infty}^{\infty} e^{vt_0} p_1(v)\, dv.\end{aligned} \tag{11}$$

The characteristic function for $\hat{p}_1(v)$,

$$Ee^{ixv} = g(u_0 + x)/g(u_0), \tag{12}$$

will be used in Section 4.

3. Remarks concerning the asymptotic series for $Q_N(y)$

(a) *General values of n.* A_n can be calculated for general values of n by using a recurrence relation for the coefficients in the classical asymptotic series. Thus, from Equation (103) of [8], (changing n to j)

$$A_j = A_0 N^{-j} \sum_{n=0}^{2j} (-\mu)^{2j-n} \sum_{m=0}^{n} d_{m,n} (-2)^{m+j} (\tfrac{1}{2})_{m+j}, \tag{13}$$

where $d_{m,n}$ is computed step by step from

$$d_{m+1,n+1} = \frac{1}{n+1} \sum_{k=1}^{n-m+1} k\theta_{k+2} d_{m,n-k+1}, \qquad 0 \leqq m \leqq n \tag{14}$$

starting from $d_{00} = 1$ and $d_{0n} = 0$ for $n > 0$. For $m = 0$ and $n \geqq 1$ we get $d_{1n} = \theta_{n+2}$ and for $m = n$, $d_{nn} = \theta_3^n/n!$. It is often convenient to use $(\frac{1}{2})_{m+j} = \Gamma(m + j + \frac{1}{2})/\pi^{\frac{1}{2}}$ in (13).

Daniels' series [5] for the probability density $p_N(y) = -(d/dy)Q_N(y)$ is

$$p_N(y) \sim [2\pi N\phi^{(2)}]^{-\frac{1}{2}} \exp(f_0)\left[1 - \frac{3}{2N}(5\theta_3^2 - 2\theta_4) + \cdots\right]. \tag{15}$$

Equation (15) can be written as $p_N(y) \sim \sum \tilde{A}_j$ where

$$\begin{aligned}\tilde{A}_0 &= [2\pi N\phi^{(2)}]^{-\frac{1}{2}} \exp(f_0)\\ \tilde{A}_j &= \tilde{A}_0 N^{-j} \sum_{m=1}^{2j} d_{m,2j} (-2)^{m+j} (\tfrac{1}{2})_{m+j}\end{aligned} \tag{16}$$

and $d_{m,2j}$ is given by (14). Replacing A_0 by $\tilde{A}_0$ and setting $\mu = 0$ in the expression (13) for A_j gives $\tilde{A}_j$. For example, $\tilde{A}_1$ and $\tilde{A}_2$ can be obtained from (9) by setting $\mu = 0$ in A_1 and A_2.

(b) *The case* $y = \bar{y}$. When $y = \bar{y}$, t_0 is zero, and A_n and B_n are infinite, but the limit of $A_n - B_n$ remains finite as $y \to \bar{y}$. In this case a classical saddle point analysis gives

$$
\begin{aligned}
Q_N(\bar{y}) \sim \tfrac{1}{2} + (2\pi N)^{-\frac{1}{2}}[&-\theta_3 + N^{-1}(\tfrac{35}{2}\theta_3^3 - 15\theta_3\theta_4 + 3\theta_5) \\
&-\tfrac{15}{2}N^{-2}(\tfrac{3003}{20}\theta_3^5 - 231\theta_3^3\theta_4 + 63\theta_3\theta_4^2 + 63\theta_3^2\theta_5 \\
&-14\theta_4\theta_5 - 14\theta_3\theta_6 + 2\theta_7) + \cdots].
\end{aligned} \tag{17}
$$

It can be verified that the terms in (17) agree with corresponding ones in the Edgeworth series ([1], No. 26.2.48) for the case $y = \bar{y}$.

(c) *Another form of* (3). The form (3) of the asymptotic series for $Q_N(y)$ is convenient for calculations when N is fixed and y varies. A different form, useful in analyzing the errors, can be obtained by introducing C_n defined by

$$A_n - B_n = C_n N^{-n-\frac{1}{2}} \exp(N\gamma_0)\Gamma(n+\tfrac{1}{2})/\pi, \tag{18}$$

where $\gamma_0 = \gamma(u_0) = f_0/N$ and

$$\gamma(u) = \phi(iu) - iur. \tag{19}$$

Then (3) becomes

$$Q_N(y) \sim \tfrac{1}{2}\operatorname{erfc}(\sqrt{-N\gamma_0}) + \frac{1}{\pi} e^{N\gamma_0} \sum_{n=0}^{\infty} C_n \Gamma(n+\tfrac{1}{2}) N^{-n-\frac{1}{2}}. \tag{20}$$

The structure of $A_n - B_n$ shows that C_n does not depend explicitly on N although there is an implicit dependence via r.

(d) *Integration of the Daniels series.* An interesting question arises regarding the accuracy of our asymptotic series for $Q_N(y)$ compared with that of the series obtained by integrating Daniels' series (15) for $p_N(y)$. It appears difficult to give an answer in the general case because of the complexity of the integration. However, some insight can be obtained by examining the exponential distribution discussed later in Section 6. From the results given there it can be shown that the Daniels series is

$$p_N(y) \sim p_{N\,\mathrm{ex}}(y)[N!/N^N e^{-N}\sqrt{2\pi N}]\left(1 - \frac{1}{12N} + \frac{1}{288N^2} + \cdots\right),$$

where $p_{N\,\mathrm{ex}}(y)$ is the exact density $y^{N-1}e^{-y}/(N-1)!$, $y > 0$. When we integrate and take the first two terms, for example, we get

$$\hat{Q}A(2) = Q_{\mathrm{ex}}[N!/N^N e^{-N}\sqrt{2\pi N}]\left(1 - \frac{1}{12N}\right),$$

where Q_{ex} is the exact value (38) of $Q_N(y)$. $\hat{Q}A(2)$ is to be compared with our

$$QA(2) = \tfrac{1}{2}\operatorname{erfc}(\sqrt{-f_0}) + (A_0 - B_0) + (A_1 - B_1)$$

given by (45). The 'relative error' $\mathscr{E}$ of $QA(2)$ is plotted in Figure 1 for the case $N=5$. For $N=5$ the relative error of $\hat{Q}A(2)$ can be shown to be $\hat{\mathscr{E}} = -0{\cdot}00016$ when $y>5$. Comparison with Figure 1 shows that $\mathscr{E}$ and $\hat{\mathscr{E}}$ are of the same order of magnitude in this particular case.

This example suggests the conjecture that integration of Daniels' series and our asymptotic series for $Q_N(y)$ both give approximations to $Q_N(y)$ that are in error by the same order of magnitude.

(e) *Existence of* u_0. We conclude this section with some remarks regarding the existence of u_0 that are based on Daniels' work [5]. Let $p_1(v)$ be zero outside of $a \leqq v \leqq b$ where a or b may be infinite. It can be shown that u_0 exists for every value of y/N between a and b if the integral

$$\int_a^b e^{tv} p_1(v)\, dv$$

exists for all real values of t. Examples are:

(i) Finite a and b.

(ii) $a=-\infty$, $b=\infty$ and $p_1(v) \leqq A \exp(-|v|^{1+\varepsilon})$ where $\varepsilon>0$ and A is a constant.

When the integral does not exist for all values of t, the question becomes more complicated:

(iii) For $a=0$, $b=\infty$, $p_1(v) = A\exp(-v^{1-\varepsilon})$ and $0<\varepsilon<1$, u_0 exists when $0 \leqq y \leqq \bar{y}$, but not when $y>\bar{y}$.

(iv) When $a=0$, $b=\infty$, $p_1(v) = Av^{\alpha-1}(1+v)^{-\beta}e^{-v}$ and $\alpha>0$, u_0 exists for the entire range $0\leqq y \leqq \infty$ if $\beta \leqq \alpha+1$. When $\beta>\alpha+1$, u_0 exists only if $0 \leqq y/N \leqq 1/(\beta-\alpha-1)$. In both cases u_0 runs from $i\infty$ to $-i$ as y runs over the range for which u_0 exists.

(v) When $p_1(v) = \tfrac{1}{2}\exp(-|v|)$ there are two saddle points on the imaginary axis. As y runs from $-\infty$ to $+\infty$, the principal saddle point u_0 runs from $+i$ to $-i$.

4. Sufficient conditions for the series to be asymptotic

Here sufficient conditions are given for the series (3) to be asymptotic in $1/N$ when $r=y/N$ is fixed. We consider only the case $y>\bar{y}$ in detail. The analysis for the case $y<\bar{y}$ is very similar and will not be repeated. For $y>\bar{y}$, t_0 is positive and $u_0=-it_0$ lies on the negative imaginary u-axis.

Suppose that u_0 has been determined for some fixed value of r, and let P denote the straight line path $\operatorname{Im}(u-u_0)=0$ joining $u_0 \pm \infty$ and passing through u_0. Let the characteristic function $g(u)$ of $p_1(v)$ satisfy the conditions:

(i) $g(u)$ is analytic throughout a strip $-t_0-\varepsilon \leqq \mathrm{Im}(u) \leqq \varepsilon$ where ε is some positive constant.

(ii) Positive constants α, c_0 and c_1 exist such that $|g(u)|<c_0/|u|^{\alpha}$ when $|u|>c_1$ on the path P.

These conditions allow us to displace the path of integration in the integral (2) for $Q_N(y)$ down to P. After making this displacement, the expression (4) for f_0, and the relation $g(u)=\exp[\phi(iu)]$ are used to rewrite (2) as

$$(21)\quad Q_N(y)=\frac{\exp(f_0)}{\pi}\mathrm{Re}\int_0^{\infty}\exp\{N[\phi(iu_0+ix)-\phi(iu_0)-ixr]\}\,dx/(iu_0+ix),$$

where $x=u-u_0$. In (21)

$$(22)\qquad \exp[\phi(iu_0+ix)-\phi(iu_0)]=g(u_0+x)/g(u_0),$$

where, from (12), the right-hand side of (22) is the characteristic function $E\exp(ivx)$ of the associated density $\hat{p}_1(v)$. Since x is real in (21), $|g(u_0+x)/g(u_0)|\leqq 1$ with equality only at $x=0$ because $\hat{p}_1(v)$ has no lattice component (as a consequence of (ii)). Therefore in (21)

$$\mathrm{Re}[\phi(iu_0+ix)-\phi(iu_0)-ixr]\leqq 0$$

with equality only at $x=0$. Furthermore, Condition (i) shows that $\phi(iu_0+ix)-\phi(iu_0)-ixr$ can be expanded in a power series in x that converges in the neighborhood of $x=0$.

The preceding discussion and the fact that the contribution to $Q_N(y)$ of the region around u_0 is $\sum A_n$ (as already mentioned in connection with (9)) shows that Conditions (i) and (ii) are sufficient to guarantee that

$$(23)\qquad Q_N(y)\sim\sum_{n=0}^{\infty}A_n$$

as $N\to\infty$ (see Olver [6], Chapter 4, Section 6). Subtracting the known asymptotic series

$$(24)\qquad \tfrac{1}{2}\,\mathrm{erfc}\,(\sqrt{-f_0})\sim\sum_{n=0}^{\infty}B_n$$

completes the proof that (3) is indeed an asymptotic series for $Q_N(y)$ when $y>\bar{y}$ and $r=y/N$ is fixed.

5. Error analysis

Let the conditions of Section 4 be satisfied and rewrite (2) as

$$(25)\qquad Q_N(y)=\mathrm{Re}\frac{1}{\pi}\int_{u_0}^{\infty+i0}e^{N\gamma(u)}\,du/(iu),$$

where $\gamma(u)=\phi(iu)-iur$. Let the right-hand branch of the steepest descent path of $\exp[\gamma(u)]$ from u_0 end at the sink $u=s_0$ where $\gamma(s_0)=-\infty$. Then

$$Q_N(y)=I(u_0, s_0)+I(s_0, \infty+i0), \tag{26}$$

where $I(u_1, u_2)$ denotes the integral in (25) with limits u_1, u_2.

The sink s_0 occurs at a zero of the characteristic function $g(u)$, but it may shift from one zero to another as $r=y/N$ changes. The path of steepest descent from u_0 is given by $\operatorname{Im}[\gamma(u_0)-\gamma(u)]=0$. If the path cannot be determined easily by analysis it can be traced step by step by starting at $u_0+\Delta$ and using

$$u_{l+1}=u_l-|\gamma'(u)|\,\Delta/\gamma'(u), \tag{27}$$

where Δ is the step length and $\gamma'(u)=d\gamma(u)/du$.

It is convenient to regard the right-hand side of (3) as the asymptotic expansion of $I(u_0, s_0)$ and $I(s_0, \infty+i0)$ as an exponentially small correction term. This point of view is helpful in explaining the fact that an asymptotic series may sometimes appear to be more accurate than it actually is (Olver [6], p. 95).

Let $m\geqq 1$ and define the partial sum $QA(m)$ by

$$QA(m)\triangleq\tfrac{1}{2}\operatorname{erfc}(\sqrt{-f_0})+\sum_{n=0}^{m-1}(A_n-B_n). \tag{28}$$

Then the error in $QA(m)$ is

$$\begin{aligned} QA(m)-Q_N(y)&=[QA(m)-I(u_0, s_0)]+[-I(s_0, \infty+i0)]\\ &=EP_m+ES, \end{aligned} \tag{29}$$

where the principal error EP_m and the exponentially small error ES are defined by the quantities within the brackets:

$$\begin{aligned} EP_m&\triangleq QA(m)-\operatorname{Re}\frac{1}{\pi}\int_{u_0}^{s_0}e^{N\gamma(u)}\,du/(iu),\\ ES&\triangleq-\operatorname{Re}\frac{1}{\pi}\int_{s_0}^{\infty+i0}e^{N\gamma(u)}\,du/(iu). \end{aligned} \tag{30}$$

ES depends only on N. It can be zero in some cases and greater than EP_m in others. A bound for $|EP_m|$ can be obtained by modifying Olver's error bound ([6], p. 89) for asymptotic series so as to take $\operatorname{erfc}(\sqrt{-f_0})$ into account. If (A_m-B_m) is not zero it can be shown that

$$|EP_m|\leqq[N/(N-\sigma_m)]^{m+\frac{1}{2}}\,|A_m-B_m|, \tag{31}$$

where σ_m is the supremum of a complicated function of m, r and u along the steepest path from u_0 to s_0:

$$
(32)\qquad \begin{aligned}
&\sigma_m = \sup_{\tau\in(0,\infty)} [\tau^{-1}\ln|F_m(\tau)/(C_m\tau^{m-\frac{1}{2}})|],\\
&\tau = \gamma_0 - \gamma(u),\\
&F_m(\tau) = \tfrac{1}{2}(-\gamma_0\tau)^{-\frac{1}{2}}(1-\tau/\gamma_0)^{-1} - \operatorname{Re}\left(\frac{1}{iu}\frac{du}{d\tau}\right) + \sum_{n=0}^{m-1} C_n\tau^{n-\frac{1}{2}}.
\end{aligned}
$$

Here C_n is defined by (18) and $\gamma_0 = \gamma(u_0)$.

6. Example—The exponential distribution

(a) *Asymptotic series.* The characteristic function for the one-sided density $p_1(v) = e^{-v}$, $v > 0$ is

$$
(33)\qquad \begin{aligned}
g(u) &= \int_0^\infty e^{iuv-v}\,dv\\
&= 1/(1-iu)
\end{aligned}
$$

and

$$
(34)\qquad \phi(iu) = \ln g(u) = -\ln(1-iu).
$$

The saddle point equation (6) gives $u_0 = -it_0$ where $t_0 = 1 - r^{-1}$ with $r = y/N$. The quantities needed to calculate A_n and B_n are

$$
(35)\qquad \begin{aligned}
&f_0 = N(\ln r - r + 1), \qquad \phi^{(n)} = (n-1)!\, r^n.\\
&\theta_n = 1/n, \qquad \mu = 1/(r-1),
\end{aligned}
$$

and (8) and (9) give

$$
(36)\qquad \begin{aligned}
&A_0 = \mu(2\pi N)^{-\frac{1}{2}} e^{f_0} && B_0 = \tfrac{1}{2}(-\pi f_0)^{-\frac{1}{2}} e^{f_0},\\
&A_1 = A_0 N^{-1}(-\mu^2 - \mu - \tfrac{1}{12}), && B_1 = B_0/(2f_0),\\
&A_2 = A_0 N^{-2}(3\mu^4 + 5\mu^3 + \tfrac{25}{12}\mu^2 + \tfrac{1}{12}\mu + \tfrac{1}{288}), && B_2 = 3B_1/(2f_0).
\end{aligned}
$$

When $y = \bar{y} = N$, r is 1, μ is infinite, and it is necessary to use the Edgeworth series

$$
(37)\qquad Q_N(y) \sim \tfrac{1}{2} + (2\pi N)^{-\frac{1}{2}}\left[-\frac{1}{3} - \frac{1}{N}\frac{1}{540} + \frac{1}{N^2}\frac{25}{6048} + \cdots\right].
$$

The exact expression for $Q_N(y)$ is

$$
(38)\qquad Q_N(y) = e^{-y}\sum_{n=0}^{N-1} y^n/n!.
$$

(b) *Error analysis.* An examination of the path of steepest descent of $\exp[N\gamma(u)]$, where $\gamma(u) = -\ln(1-iu) - iur$, from u_0 shows that the sink s_0 occurs at $s_0 = \pi/r - i\infty$, and consequently the exponentially small error ES is zero. To illustrate the bound (31) for EP_m consider the case $r = y/N = 2$ and $m = 1$. Numerical evaluation of (32) at points along the path between u_0 and s_0 shows that the supremum occurs at $\tau = 6{\cdot}94$. It has the value $\sigma = 0{\cdot}085$ and (31) becomes

$$|EP_1| \leqq [N/(N-0{\cdot}085)]^{\frac{3}{2}}\,|A_1 - B_1|. \tag{39}$$

Calculations for $N = 5$ and $y = rN = 10$ show that

$$\begin{aligned} Q_5(10) &= 0{\cdot}02925\cdots, \\ EP_1 &= QA(1) - Q_5(10) = -2{\cdot}18(-5), \\ A_1 - B_1 &= -2{\cdot}58(-5). \end{aligned}$$

Therefore (39) gives $|EP_1| \leqq 2{\cdot}65(-5)$ which is slightly larger than $|EP_1| = 2{\cdot}18(-5)$.

7. Example—The uniform distribution

The characteristic function corresponding to $p_1(v) = \frac{1}{2}$ for $-1 < v < 1$ is $g(u) = \sin u/u$, and $\phi(iu) = \ln g(u) = \ln(\sin u/u)$. In order to calculate (3) we need the values of

$$\phi(t) = \ln(\sinh t/t) \tag{40}$$

and its derivatives at the real root t_0 of

$$\cosh t - t^{-1} - r = 0. \tag{41}$$

Equation (41) can be solved by starting with $t_0 \approx 1/(1-r)$ and using the Newton–Raphson method. Differentiation of (40) gives

$$\begin{aligned} (d/dt)^2\phi(t) &= -\operatorname{csch}^2 t + t^{-2} \\ (d/dt)^3\phi(t) &= (2\operatorname{csch}^3 t)\cosh t - 2t^{-3} \end{aligned}$$

and so on. For t real we also have

$$(d/dt)^n\phi(t) = (-)^{n-1}(n-1)!\,2\sum_{l=1}^{\infty}\operatorname{Re}(t+i\pi l)^{-n}$$

which is useful when n is large and the recursion relation (13) is used to calculate A_n. For the error calculations the exact expression

$$Q_N(y) = \frac{1}{2^N N!}\sum_{k=0}^{K}(-)^k\binom{N}{k}(N-y-2k)^N \tag{42}$$

was used. Here K is the largest integer in $(N-y)/2$.

A trial calculation was made using $N=4$ and $y=3\cdot2$. In this case the smallest term in (3) is $A_3-B_3=-2\cdot8(-6)$ and the error made in stopping with this term is $QA(4)-Q_4(3\cdot2)=4\cdot9(-6)$, i.e.,

(43) $$EP_4+ES=4\cdot9(-6).$$

This is small compared to the exact value $Q_4(3\cdot2)=1066\cdot7(-6)$.

(a) *The error ES.* For the uniform distribution, the function $\gamma(u)$ appearing in the definition (30) of ES is $\gamma(u)=\ln(\sin u/u)-iur$. When the path of steepest descent from u_0 is calculated by using $\gamma'(u)=\cot u-u^{-1}-ir$ in the step-by-step formula (27) it is found that s_0 is one of the zeros, say $l\pi$, of $\sin u/u$. If r is between 0 and a number slightly larger than $0\cdot7$, l is 1. For $r=0\cdot8$, l is 2, and for $r=0\cdot9$, l is 5. From (30) and $r=y/N$,

(44) $$ES=\frac{1}{\pi}\int_{l\pi}^{\infty}\left(\frac{\sin u}{u}\right)^N\frac{\sin uy}{u}\,du.$$

For $N=4$, $y=3\cdot2$ we get $r=0\cdot8$, $l=2$ and the value of ES calculated from (44) is $-2\cdot1(-6)$. This is an appreciable fraction of the total error $4\cdot9(-6)$ stated above for our partial sum $QA(4)$. Incidentally, when $l=0$ in (44) the integral is equal to $Q_N(y)$.

8. Comparison with other approximations

Here we compare our first two partial sums

(45) $$\begin{aligned} QA(1)&=\tfrac{1}{2}\operatorname{erfc}(\sqrt{-f_0})+(A_0-B_0),\\ QA(2)&=QA(1)+(A_1-B_1), \end{aligned}$$

with other approximations to $Q_N(y)$. The comparison is based on the relative error defined by

(46) $$\begin{aligned} \mathscr{E}&=(Q_{ap}-Q_{ex})/Q_{ex}, && y>\bar{y},\\ \mathscr{E}&=[(1-Q_{ap})-(1-Q_{ex})]/(1-Q_{ex}), && y<\bar{y}, \end{aligned}$$

where Q_{ap} is the approximation and Q_{ex} is the exact value $Q_N(y)$.

Figure 1 shows $|\mathscr{E}|$ for approximations to $Q_5(y)$ when the individual density is $p_1(v)=\exp(-v)$, $v>0$ (Section 6). The sign of $\mathscr{E}$ is indicated by the + or − on the curve. A portion of $QA(2)$ around $y=\bar{y}=5$ has been omitted to reduce the clutter. The 'Edge$_2$' curve is calculated from an Edgeworth series ([1], No. 26.2.48) in which the last term depends on the cumulants $\kappa_2=\sigma^2=1$, $\kappa_3=2$, $\kappa_4=6$ of v and the third and fifth derivatives of $\exp(-x^2/2)$ where

(47) $$\begin{aligned} x&=(y-\bar{y})/(\sigma\sqrt{N}),\\ &=(y-5)/2\cdot236, \qquad N=5. \end{aligned}$$

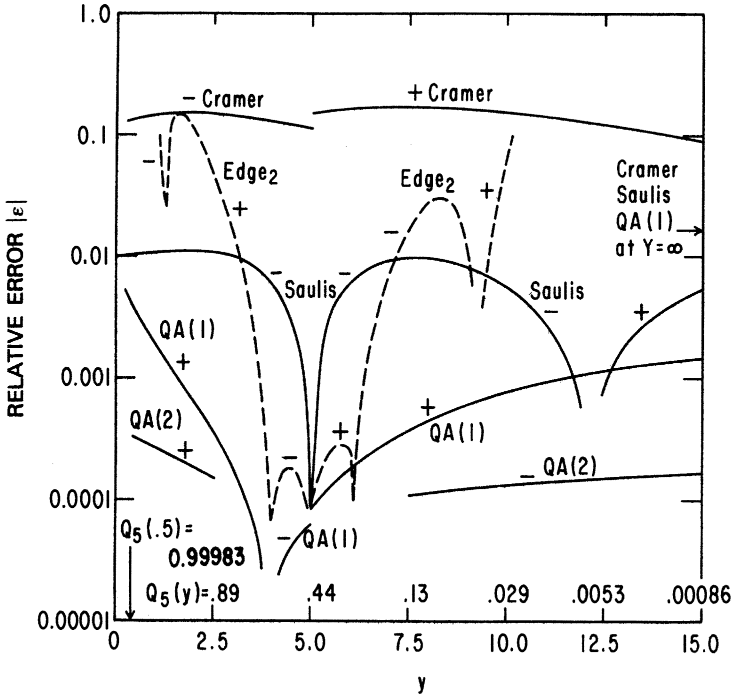

Figure 1
Relative error of various approximations to $Q_5(y)$ when $p_1(v)=\exp(-v)$, $v>0$.

For $y>\bar{y}$ the 'Cramér' curve is calculated from the approximation

$$Q_N(y)\approx[1-\Phi(x)]\exp[f_0+\tfrac{1}{2}x^2] \tag{48}$$

obtained by deleting the 'order of' term in the equation for $Q_N(y)$ presented in [7], p. 219. Here $1-\Phi(x)=\frac{1}{2}\operatorname{erfc}(x/\sqrt{2})$ and x is given by (47). A similar result holds when $y<\bar{y}$.

In (48) we have made use of the fact that Cramér's function $\lambda(z)$ is related to our f_0 by

$$(x^3/\sqrt{N})\lambda(x/\sqrt{N})=f_0+\tfrac{1}{2}x^2. \tag{49}$$

The 'Saulis' curve for $y>\bar{y}$ is calculated in much the same way as the 'Cramér' curve by using an expression given by Petrov ([7], p. 249) which represents the first two terms in a general series given by Saulis [10].

Figure 2 shows $|\mathscr{E}|$ for approximations to $Q_5(y)$ when $p_1(y)=\frac{1}{2}$, $|v|<1$, the uniform distribution (Section 7). The Edge$_2$ curve is calculated from the same general formulas as for the exponential distribution but now the cumulants are $\kappa_2=\sigma^2=\frac{1}{3}$, $\kappa_3=0$, $\kappa_4=-\frac{2}{15}$ and $x=y/(\sigma\sqrt{N})=y\sqrt{\frac{3}{5}}$. It turns out that, because $\kappa_3=0$, the two-term Saulis formula reduces to Cramér's result.

It is seen that $QA(1)$ and $QA(2)$ do quite well over the entire range of y.

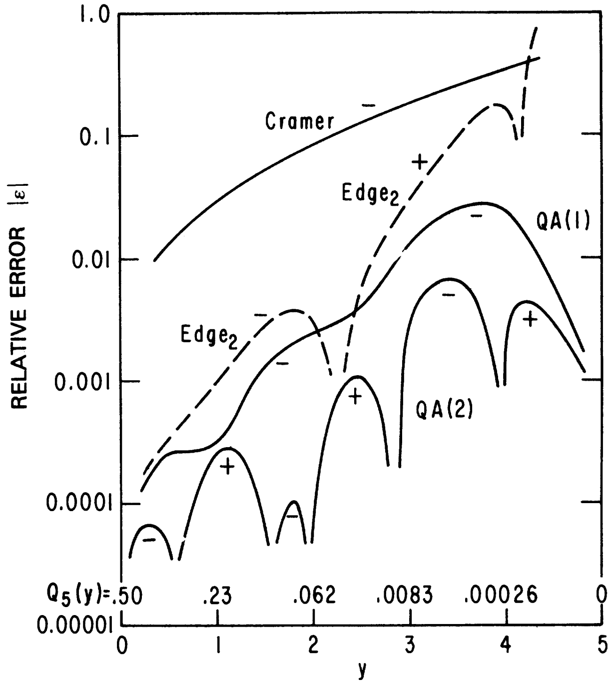

Figure 2
Relative error of various approximations to $Q_5(y)$ when $p_1(v)=\frac{1}{2}$, $|v|<1$.

Appendix. Uniform asymptotic series for integrals

The integral (2) for $Q_N(y)$ is a special case (with a change of notation) of the integral

$$J=\int_{L'} t^{\lambda-1}g(t)e^{xh(t)}\,dt, \tag{50}$$

where x is large and positive. Methods associated with the names of Bleistein and Ursell for expanding (50) in a 'uniform asymptotic series' have been discussed in [8].

(a) *General comments.* First assume that the term $\exp[xh(t)]$ in (50) has μ simple saddle points (where the first derivative of $h(t)$ vanishes but the second does not) and that λ is not a positive integer. The saddle points and the origin lie within a relatively small 'critical' region in the t-plane through which the path L' passes. The functions $g(t)$ and $h(t)$ are analytic throughout the critical region, $g(0)\neq 0$, $h(0)=0$ and $t=0$ is not a saddle point.

In the critical region $h(t)$ behaves like a polynomial of degree $\mu+1$. Let v be a new variable such that $F(v)=h(t)$ where $F(v)$ is a polynomial of degree $\mu+1$ in v, and v is nearly proportional to t in the critical region. The choice of $F(t)$ is discussed in [8]. This change of variable carries (50) into

$$J = \int_L v^{\lambda-1} f(v) e^{xF(v)}\, dv, \tag{51}$$

where L in the v-plane corresponds to L' in the t-plane. Let $t_1, t_2, \cdots, t_\mu$ be the saddle points in the t-plane and $v_1, v_2, \cdots, v_\mu$ the corresponding ones in the v-plane. Deforming L into paths of steepest descent and considering the separate contributions of the saddle points leads to an asymptotic expansion of the form

$$J \sim \sum_{l=0}^{\mu} V_l(x)(p_{0l} + p_{1l}x^{-1} + p_{2l}x^{-2} + \cdots), \tag{52}$$

$$V_l(x) = \int_L v^{l+\lambda-1} e^{xF(v)}\, dv, \tag{53}$$

where $V_l(x)$ is regarded as a tabulated or easily computed function. It turns out that p_{nl} does not depend on L' or L, and by choosing suitable paths we can get a set of equations that can be solved for the p_{nl}'s.

(b) *The one-saddle-point case.* For illustration consider the case $\mu = 1$. A treatment of this case which differs somewhat from the following one is given in Appendix F of [8], and entirely different treatments are given by Bleistein in Sections 6 and 7 of [2] and by van der Waerden [11]. See also the excellent discussion of uniform asymptotic expansions given in Chapter 9 of Bleistein and Handelsman [3]. We seek an expansion of the form (52) with $\mu = 1$ when $V_0(x)$ and $V_1(x)$ are given by (53) in which $F(v)$ is a second-degree polynomial. A convenient choice is

$$F(v) = v^2 - 2v_1 v. \tag{54}$$

Let L_1' be the path of steepest descent that has t_1 as its highest point. A classical saddle-point expansion about t_1 gives

$$\begin{aligned} J_1 &= \int_{L_1'} t^{\lambda-1} g(t) e^{xh(t)}\, dt \\ &\sim \exp[xh(t_1)] \sum_{n=0}^{\infty} \alpha_{1n} x^{-n-\frac{1}{2}}, \end{aligned} \tag{55}$$

where the α_{1n}'s can be calculated from the derivatives of $g(t)$ and $h(t)$ at t_1 (see, for example, (103) of [8]). Similarly, from (52) and (53), J_1 can also be expressed as

$$J_1 \sim \sum_{l=0}^{1} [V_l(x)]_1 (p_{0l} + p_{1l}x^{-1} + \cdots), \tag{56}$$

(57)
$$[V_l(x)]_1 = \int_{L_1} v^{l+\lambda-1} e^{xF(v)}\, dv \sim \exp[xF(v_1)] \sum_{m=0}^{\infty} \beta_{1lm} x^{-m-\frac{1}{2}},$$

where L_1 is the path that has v_1 as its highest point.

Putting (57) in (56), equating coefficients of $x^{-n-\frac{1}{2}}$ in (55) to those in (56), and using

(58)
$$\beta_{110} = \beta_{100} v_1$$

gives a set of relations, the nth of which is

(59)
$$p_{n1} = -\frac{p_{n0}}{v_1} + \frac{\alpha_{1n}}{\beta_{110}} - \frac{1}{\beta_{110}} \sum_{k=0}^{n-1} (\beta_{1,0,n-k} p_{k0} + \beta_{1,1,n-k} p_{k1}),$$

where the summation is omitted when n is 0.

Another set of relations can be obtained by treating the singularity at the origin in somewhat the same way as the saddle point. If the singularity is a branch point we take L_0' and L_0 to be loops enclosing the branch cuts running out from the origins. If the singularity is a pole we take L_0' and L_0 to be small circles around the origins. In any case

(60)
$$J_0 = \int_{L_0'} t^{\lambda-1} g(t) e^{xh(t)}\, dt \sim \sum_{n=0}^{\infty} \alpha_{0n} x^{-n-\lambda},$$

(61)
$$[V_l(x)]_0 = \int_{L_0} v^{l+\lambda-1} e^{xF(v)}\, dv \sim \sum_{m=0}^{\infty} \beta_{0lm} x^{-m-l-\lambda}.$$

Equating the two asymptotic series for J_0 leads to

(62)
$$p_{n0} = \frac{\alpha_{0n}}{\beta_{000}} - \frac{1}{\beta_{000}} \sum_{k=0}^{n-1} (\beta_{0,0,n-k} p_{k0} + \beta_{0,1,n-k} p_{k1}).$$

Equations (59) and (62) can be used to calculate p_{n0} and p_{n1} step by step, starting with $p_{00} = \alpha_{00}/\beta_{000}$ from (62).

(c) *Application to* $Q_N(y)$. For $Q_N(y)$ we consider the special case in which $\lambda = 0$, $g(t) = 1$, $\mu = 1$, t_1 is real, and L' runs from $-i\infty$ to $+i\infty$ with an indentation to the right at $t = 0$ (see Section 9 of [8] and Bleistein [2]). When $F(v)$ is defined by (54), (53) gives

(63)
$$V_0(x) = i\pi \operatorname{erfc}(v_1 x^{\frac{1}{2}}), \qquad V_1(x) = i(\pi/x)^{\frac{1}{2}} \exp(-xv_1^2).$$

If L_0' and L_0 are taken to be small circles about the origin,

$$J_0 = 2\pi i, \qquad [V_0(x)]_0 = 2\pi i, \qquad [V_1(x)]_0 = 0,$$

and consequently all of the α_{0n}, β_{00m}, β_{01m} are zero except $\alpha_{00} = 2\pi i$ and

$\beta_{000}=2\pi i$. The recurrence relation (62) then gives

(64) $$p_{00}=1, \qquad p_{n0}=0, \qquad n \geqq 1.$$

The α_{1n}'s are defined by

(65) $$\int_{L_1'} t^{-1} e^{xh(t)}\,dt \sim e^{xh(t_1)} \sum_{n=0}^{\infty} \alpha_{1n} x^{-n-\frac{1}{2}},$$

where L_1' runs upward through t_1. Taking L_1 to run from $-i\infty$ to $+i\infty$ through v_1 makes $[V_l(x)]_1 = V_l(x)$, $l=0$, 1. Noting that the asymptotic series for $V_1(x)$ consists of only the leading term shows that β_{11m} in (57) is 0 except for $\beta_{110}=i\pi^{\frac{1}{2}}$. Similarly, the asymptotic series for $V_0(x)$ leads to

(66) $$\beta_{10m} = i\pi^{\frac{1}{2}}(-)^m(\tfrac{1}{2})_m v_1^{-2m-1}, \qquad m=0,1,2\cdots.$$

Note that $\beta_{110}=v_1\beta_{100}$ as it should according to (58). The recurrence relation (59) then gives

(67) $$p_{n1}=(\alpha_{1n}-\beta_{10n})/\beta_{110}.$$

Inserting the values of p_{n0} and p_{n1} in the series (52) leads to

(68) $$\int_{L'} t^{-1}e^{xh(t)}\,dt \sim i\pi \operatorname{erfc}(v_1 x^{\frac{1}{2}}) + \exp(-xv_1^2) \sum_{n=0}^{\infty} (\alpha_{1n}-\beta_{10n})x^{-n-\frac{1}{2}}.$$

Since $h(t_1)=F(v_1)=-v_1^2$ this series has essentially the same form as the series (3) for $Q_N(y)$.

References

[1] Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions.* Dover, New York.

[2] Bleistein, N. (1966) Uniform asymptotic expansions of integrals with stationary point near algebraic singularity. *Comm. Pure Appl. Math.* **19,** 353–370.

[3] Bleistein, N. and Handelsman, R. A. (1975) *Asymptotic Expansion of Integrals.* Holt, Rinehart and Winston, New York.

[4] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23,** 493–507.

[5] Daniels, H. E. (1954) Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25,** 631–650.

[6] Olver, F. W. J. (1974) *Asymptotics and Special Functions.* Academic Press, New York.

[7] Petrov, V. V. (1973) *Sums of Independent Random Variables,* trans. A. A. Brown. Springer-Verlag, New York.

[8] Rice, S. O. (1968) Uniform asymptotic expansions for saddle point integrals—Applications to a probability distribution occurring in noise theory. *Bell Syst. Tech. J.* **47,** 1971–2013.

[9] Roberts, J. B. (1972) Distribution of the response of linear systems to Poisson distributed random impulses. *J. Sound and Vibration* **24,** 23–34.

[10] Saulis, L. (1969) Asymptotic expansions of probabilities of large deviations (in Russian). *Litovsk. Mat. Sb.* **9,** 605–625.

[11] Van Der Waerden, B. L. (1951) On the method of saddle points. *Appl. Sci. Res.* B**2,** 33–45.