SADDLEPOINT APPROXIMATIONS IN STATISTICS[1]

By H. E. Daniels

University of Cambridge and University of Chicago

1. Introduction and summary. It is often required to approximate to the distribution of some statistic whose exact distribution cannot be conveniently obtained. When the first few moments are known, a common procedure is to fit a law of the Pearson or Edgeworth type having the same moments as far as they are given. Both these methods are often satisfactory in practice, but have the drawback that errors in the "tail" regions of the distribution are sometimes comparable with the frequencies themselves. The Edgeworth approximation in particular notoriously can assume negative values in such regions.

The characteristic function of the statistic may be known, and the difficulty is then the analytical one of inverting a Fourier transform explicitly. In this paper we show that for a statistic such as the mean of a sample of size n, or the ratio of two such means, a satisfactory approximation to its probability density, when it exists, can be obtained nearly always by the method of steepest descents. This gives an asymptotic expansion in powers of n^{-1} whose dominant term, called the saddlepoint approximation, has a number of desirable features. The error incurred by its use is $O(n^{-1})$ as against the more usual $O(n^{-1/2})$ associated with the normal approximation. Moreover it is shown that in an important class of cases the *relative* error of the approximation is uniformly $O(n^{-1})$ over the whole admissible range of the variable.

The method of steepest descents was first used systematically by Debye for Bessel functions of large order (Watson [17]) and was introduced by Darwin and Fowler (Fowler [9]) into statistical mechanics, where it has remained an indispensable tool. Apart from the work of Jeffreys [12] and occasional isolated applications by other writers (e.g. Cox [2]), the technique has been largely ignored by writers on statistical theory.

In the present paper, distributions having probability densities are discussed first, the saddlepoint approximation and its associated asymptotic expansion being obtained for the probability density of the mean $\bar{x}$ of a sample of n. It is shown how the steepest descents technique is related to an alternative method used by Khinchin [14] and, in a slightly different context, by Cramér [5]. General conditions are established under which the relative error of the saddlepoint approximation is $O(n^{-1})$ uniformly for all admissible $\bar{x}$, with a corresponding result for the asymptotic expansion. The case of discrete variables is briefly discussed, and finally the method is used for approximating to the distribution of ratios.

2. Mean of n independent identically distributed random variables. Let x be a continuously distributed random variable with distribution function $F(x)$.

Received 1/16/53, revised 3/31/54.

[1] Research carried out partly under sponsorship of the Office of Naval Research.

Assume that a density function $f(x) = F'(x)$ exists and suppose the moment-generating function

$$M(T) = e^{K(T)} = \int_{-\infty}^{\infty} e^{Tx} f(x)\, dx$$

converges for real T in some nonvanishing interval containing the origin. Let $-c_1 < T < c_2$ be the largest such interval, where $0 \leqq c_1 \leqq \infty$ and $0 \leqq c_2 \leqq \infty$ but $c_1 + c_2 > 0$. Thus either c_1 or c_2 may be zero, though not both, and the moments need not all exist.

Consider the mean $\bar{x}$ of n independent x's. Its density function $f_n(\bar{x}) = F_n'(\bar{x})$ is given by the usual Fourier inversion formula

$$f_n(\bar{x}) = \frac{n}{2\pi} \int_{-\infty}^{\infty} M^n(it) e^{-nit\bar{x}}\, dt \tag{2.1}$$

(More generally $\int_{-\infty}^{\infty}$ may be replaced by $\lim_{t\to\infty} \int_{-t}^{t}$, but the argument is unaffected.) It is convenient here to employ the equivalent inversion formula

$$f_n(\bar{x}) = \frac{n}{2\pi i} \int_{\tau-i\infty}^{\tau'+i\infty} e^{n[K(T)-T\bar{x}]}\, dT \tag{2.2}$$

where $-c_1 < \mathcal{R}(T) < c_2$ on the path of integration, and $K(T)$ is the cumulant-generating function.

When n is large, an approximation to $f_n(\bar{x})$ is found by choosing the path of integration to pass through a saddlepoint of the integrand in such a way that the integrand is negligible outside its immediate neighbourhood. The saddlepoints are situated where the exponent has zero derivative, that is where

$$K'(T) = \bar{x}. \tag{2.3}$$

We shall prove in Section 5 that under general conditions (2.3) has a single real root T_0 in $(-c_1, c_2)$ for every value of $\bar{x}$ such that $0 < F_n(\bar{x}) < 1$, and that $K''(T_0) > 0$. Let us choose the path of integration to be a straight line through T_0 parallel to the imaginary axis. Since $K(T) - T\bar{x}$ has a minimum at T_0 for real T, the modulus of the integrand must have a maximum at T_0 on the chosen path. Now we can show by a familiar argument (cf. Wintner [18], p. 14) that on *any* admissible straight line parallel to the imaginary axis the integrand attains its maximum modulus only where the line crosses the real axis. For on the line $T = \tau + iy$,

$$|M(T)e^{-T\bar{x}}| = e^{-\tau\bar{x}} \left| \int_{-\infty}^{\infty} e^{(\tau+iy)x}\, dF(x) \right|$$
$$\leqq e^{-\tau\bar{x}} M(\tau).$$

Equality cannot hold for some $y \neq 0$, otherwise $\int_{-\infty}^{\infty} e^{(\tau+iy)x}\, dF(x) = M(\tau)e^{i\alpha}$

so that $\int_{-\infty}^{\infty} e^{\tau x}[1 - \cos(yx - \alpha)]\, dF(x) = 0$, which contradicts the existence of a density function. Moreover, since $M(\tau + iy) = 0(|y|^{-1})$ for large $|y|$ by the Riemann Lebesgue lemma, the integrand cannot approach arbitrarily near its maximum modulus as $|y|$ becomes large. Consequently, for the particular path chosen, only the nieghbourhood of T_0 need be considered when n is large.

The argument then proceeds formally as follows. On the contour near T_0,

$$K(T) - T\bar{x} = K(T_0) - T_0\bar{x} - \tfrac{1}{2}K''(T_0)y^2 - \tfrac{1}{6}K'''(T_0)iy^3 + \tfrac{1}{24}\cdot K^{iv}(T_0)y^4 + \cdots. \tag{2.4}$$

Setting $y = v/[nK''(T_0)]^{1/2}$ and expanding the integrand we get

$$f_n(\bar{x}) \sim \frac{1}{2\pi}\left[\frac{n}{K''(T_0)}\right]^{1/2} e^{n[K(T_0) - T_0\bar{x}]} \cdot \int_{-\infty}^{\infty} e^{-v^2/2}\left\{1 - \tfrac{1}{6}\lambda_3(T_0)\frac{iv^3}{n^{1/2}} + \frac{1}{n}[\tfrac{1}{24}\lambda_4(T_0)v^4 - \tfrac{1}{72}\lambda_3^2(T_0)v^6] + \cdots\right\} dv \tag{2.5}$$

where $\lambda_j(T) = K^{(j)}(T)/[K''(T)]^{j/2}$ for $j \geqq 3$. The odd powers of v vanish on integration and we obtain an expansion in powers of n^{-1},

$$f_n(\bar{x}) \sim g_n(\bar{x})\left\{1 + \frac{1}{n}[\tfrac{1}{8}\lambda_4(T_0) - \tfrac{5}{24}\lambda_3^2(T_0)] + \cdots\right\} \tag{2.6}$$

where $g_n(\bar{x}) = [n/2\pi K''(T_0)]^{1/2} e^{n[K(T_0) - T_0\bar{x}]}$. We call $g_n(\bar{x})$ the *saddlepoint approximation* to $f_n(\bar{x})$.

3. The method of steepest descents. It is not apparent from the above formal development that (2.6) is a proper asymptotic expansion in which the remainder is of the same order as the last term neglected. The asymptotic nature of an expansion of this type is usually established by the method of steepest descents with the aid of a lemma due to Watson [17], the path of integration being the curve of steepest descent through T_0, upon which the modulus of the integrand decreases most rapidly. An account of the method is given by Jeffreys and Jeffreys [13]. The analysis is simplified by using a "truncated" version of Watson's lemma introduced by Jeffreys and Jeffreys for this purpose.[2] The special form appropriate to the present discussion is as follows.

LEMMA. *If $\psi(z)$ is analytic in a neighbourhood of $z = 0$ and bounded for real $z = w$ in an interval $-A \leqq w \leqq B$ with $A > 0$ and $B > 0$, then*

$$\left(\frac{n}{2\pi}\right)^{1/2} \int_{-A}^{B} e^{-nw^2/2}\psi(w)\, dw \sim \psi(0) + \frac{1}{2n}\psi''(0) + \cdots + \frac{1}{(2n)^r}\frac{\psi^{(2r)}(0)}{r!} + \cdots \tag{3.1}$$

is an asymptotic expansion in powers of n^{-1}.

[2] The proof given in [13] contains an error which will be corrected in the forthcoming new edition.

To apply the lemma, deform the contour so that for $| T - T_0 | \leqq \delta$ the line $T = T_0 + iy$ is replaced by the curve of steepest descent which is that branch of $\mathcal{I}\{K(T) - T\bar{x}\} = 0$ touching $T = T_0 + iy$ at T_0, when δ is chosen small enough to exclude possible saddlepoints other than T_0. The contour is thereafter continued along the orthogonal curves of constant $\mathcal{R}\{K(T) - T\bar{x}\}$. These can easily be shown to meet the original path in points $T_0 - i\alpha$ and $T_0 + i\beta$ where $\alpha > 0$ and $\beta > 0$, if δ is small enough, since T_0 is a simple root of (2.3). The rest of the contour remains as before.

On the steepest descent curve, $K(T) - T\bar{x}$ is real and decreases steadily on each side of T_0. Make the substitution

$$
\begin{aligned}
-\tfrac{1}{2}w^2 &= K(T) - T\bar{x} - K(T_0) + T_0\bar{x} \\
&= \tfrac{1}{2}K''(T_0)(T - T_0)^2 + \tfrac{1}{6}K'''(T_0)(T - T_0)^3 + \cdots \qquad (3.2)\\
&= \tfrac{1}{2}z^2 + \tfrac{1}{6}\lambda_3(T_0)z^3 + \tfrac{1}{24}\cdot\lambda_4(T_0)z^4 + \cdots,
\end{aligned}
$$

where $z = (T - T_0)[K''(T_0)]^{1/2}$, and w is chosen to have the same sign as $\mathcal{I}(z)$ on the contour. Inversion of the series yields an expansion

$$z = iw + \tfrac{1}{6}\lambda_3(T_0)w^2 + \{\tfrac{1}{24}\cdot\lambda_4(T_0) - \tfrac{5}{72}\lambda_3^2(T_0)\}iw^3 + \cdots$$

convergent in some neighbourhood of $w = 0$. The contribution to (2.2) from this part of the contour is then

$$\frac{n}{2\pi i}\frac{e^{n[K(T_0)-T_0\bar{x}]}}{[K''(T_0)]^{1/2}}\int_{-A}^{B} e^{-nw^2/2}\frac{dz}{dw}\,dw,$$

to which Watson's lemma can be applied. Contributions to the integral from the rest of the contour are negligible since for $T = T_0 + iy$ with y outside $(-\alpha_1, \alpha_2)$ we have

$$| M(T)e^{-T\bar{x}} | \leqq \rho \, | M(T_0)e^{-T_0\bar{x}} |$$

for some $\rho < 1$, so that the extra terms contain the factor ρ^n and may be neglected. We thus obtain the asymptotic expansion

$$f_n(\bar{x}) \sim \left[\frac{n}{2\pi K''(T_0)}\right]^{1/2} e^{n[K(T_0)-T_0\bar{x}]}\left\{a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots\right\}. \qquad (3.3)$$

From the Lagrange expansion of dz/dw we find

$$a_r = \frac{1}{2^r r!}\frac{d^{2r}}{dz^{2r}}\left\{\frac{z}{iw(z)}\right\}^{2r+1}\Bigg|_{z=0}. \qquad (3.4)$$

The coefficients of this series can be shown to be identical with those obtained by the method of Section 2 (see Appendix).

4. A generalisation of the Edgeworth expansion. We now show how the work of Cramér [3], [4] on the Edgeworth series can also be employed to establish the asymptotic nature of (2.6), using a technique similar to that adopted by Cramér [5] and Khinchin [14].

It has been proved that on any admissible path of the form $T = \tau + iy$ the integrand attains its maximum modulus only at $T = \tau$. Consequently (2.6) is only one of a family of series for $f_n(\bar{x})$ which can be derived in a similar way by integrating along $T = \tau + iy$, τ taking any value in $(-c_1, c_2)$. In particular, $\tau = 0$ gives the Edgeworth series, whose asymptotic character was demonstrated by Cramér (3).

We have

$$e^{K(T)-T\bar{x}} = \int_{-\infty}^{\infty} e^{T(x-\bar{x})} f(x)\, dx = \int_{-\infty}^{\infty} e^{Tu} f(u + \bar{x})\, du. \tag{4.1}$$

On the path $T = \tau + iy$ we can put

$$e^{K(T)-T\bar{x}} = e^{K(t)-\tau x}\phi(y),$$

where

$$\phi(y) = \frac{\int_{-\infty}^{\infty} e^{iyu}\cdot e^{\tau u} f(u + \bar{x})\, du}{\int_{-\infty}^{\infty} e^{\tau u} f(u + \bar{x})\, du} \tag{4.2}$$

is the characteristic function for a random variable u having the density function $h(u) \propto e^{\tau u} f(u + \bar{x})$. The inversion formula (2.2) then becomes

$$\begin{aligned} f_n(\bar{x}) &= e^{n[K(\tau)-t\bar{x}]}\cdot(n/2\pi) \int_{-\infty}^{\infty} \phi^n(y)\, dy \\ &= e^{n[K(\tau)-\tau\bar{x}]} h_n(0) \end{aligned}$$

where $h_n(\bar{u})$ is the density function for the mean $\bar{u}$ of n independent u's. Using the fact that

$$\log \phi = [K'(\tau) - \bar{x}]iy + \sum_{j\geq 2} K^{(j)}(\tau)\, \frac{(iy)^j}{j!}$$

we may replace $h_n(0)$ by its Edgeworth series and obtain the family of asymptotic expansions

$$\begin{aligned} f_n(\bar{x}) \sim {} & \exp n\{K(\tau) - \tau\bar{x} - [K'(\tau) - \bar{x}]^2/2K''(\tau)\} \\ & \cdot [n/2\pi K''(\tau)]^{1/2}\, \{1 + A_1/n^{1/2} + A_2/n + \cdots\} \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} A_1 &= (1/3!)\lambda_3(\tau)\, H_3([K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}), \\ A_2 &= (1/4!)\lambda_4(\tau)\, H_4([K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}) \\ &\quad + (10/6!)\lambda_3^2(\tau)\, H_6([K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}), \end{aligned}$$

etc., the H's being Hermite polynomials.

When $\tau = 0$ this reduces to the Edgeworth series for $f_n(\bar{x})$. (Since c_1 or c_2 can be zero it may not be possible to take the expansion beyond a certain number of terms in this case). On the other hand when $\tau = T_0$, so that $K'(T_0) = \bar{x}$, all the odd powers of $n^{-1/2}$ vanish and we get (2.6), which is an expansion in powers of n^{-1}. In particular the dominant term $g_n(\bar{x})$ has the same accuracy as the first two terms of the Edgeworth series. Unlike the latter, however, $g_n(\bar{x})$ can never be negative, and is shown in Section 7 to have a further important advantage over the other approximations.

5. Examples. The method is applied to three examples.

EXAMPLE 5.1.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-m)^2/2\sigma^2}, \qquad -\infty \leqq x \leqq \infty.$$

$$K(T) = mT + \tfrac{1}{2}\sigma^2T^2, \qquad K'(T) = m + \sigma^2T_0 = \bar{x},$$

$$T_0 = (\bar{x} - m)/\sigma^2, \qquad K''(T_0) = \sigma^2,$$

$$g_n(\bar{x}) = \frac{1}{\sigma}\left(\frac{n}{2\pi}\right)^{1/2} e^{-n(\bar{x}-m)^2/2\sigma^2}.$$

In this case $g_n(\bar{x}) = f_n(\bar{x})$ for every value of n.

EXAMPLE 5.2. $$f(x) = (c^\alpha/\Gamma(\alpha))x^{\alpha-1}e^{-cx}, \qquad o \leq x \leq \infty.$$

$$K(T) = -\alpha \log(1 - T/c), \qquad K'(T_0) = \alpha/(c - T_0) = \bar{x},$$

$$K''(T_0) = \alpha/(c - T_0)^2 = \bar{x}^2/\alpha,$$

$$g_n(\bar{x}) = (n\alpha/2\pi)^{1/2}e^{n\alpha}(c/\alpha)^{n\alpha}\,\bar{x}^{n\alpha-1}\,e^{-nc\bar{x}}.$$

The exact result is

$$f_n(\bar{x}) = [(nc)^{n\alpha}/\Gamma(n\alpha)]\bar{x}^{n\alpha-1}e^{-nc\bar{x}}$$

which differs from $g_n(\bar{x})$ only in that $\Gamma(n\alpha)$ is replaced by Stirling's approximation in the normalising factor. As this can always be readjusted ultimately to make the total probability unity, we can regard $g_n(\bar{x})$ as being in this sense "exact" for all n.

EXAMPLE 5.3. $$f(x) = \tfrac{1}{2}, \qquad -1 \leqq x \leqq 1.$$

The density function for the mean of n independent rectangular variables in $(-1, 1)$ is known to be

$$f_n(\bar{x}) = \frac{n^n}{2^n(n-1)!}\sum_{s=0}^{n}(-1)^s\binom{n}{s}\left\langle 1 - \bar{x} - \frac{2s}{n}\right\rangle^{n-1}, \qquad |\,\bar{x}\,| \leqq 1$$

where $\langle z\rangle = z$ for $z \geqq 0$ and $=0$ for $z < 0$. (Seal [16] gives a historical note on this result.) We have

$$K(T) = \log\left(\frac{\sinh T}{T}\right), \qquad K'(T_0) = \coth T_0 - \frac{1}{T_0} = \bar{x},$$

$$K''(T_0) = \frac{1}{T_0^2} - \operatorname{cosech}^2 T_0,$$

$$g_n(\bar{x}) = \left(\frac{n}{2\pi}\right)^{1/2}\left\{\frac{1}{T_0} - \operatorname{cosech}^2 T_0\right\}^{-1/2}\left(\frac{\sinh T_0}{T_0}\right)^n e^{-T_0\bar{x}}$$

When T_0 is large and positive, $\bar{x} \sim 1 - 1/T_0$ and

$$K(T_0) \sim \log(e^{T_0}/2T_0), \qquad K''(T_0) \sim 1/T_0^2.$$

So for small $1 - \bar{x}$,

$$g_n(\bar{x}) \sim (n/2\pi)^{1/2}(\tfrac{1}{2}e)^n(1 - \bar{x})^{n-1}$$

which agrees with $f_n(\bar{x}) = [n^n/2^n(n-1)!](1-\bar{x})^{n-1}$ when $\bar{x} > 1 - 2/n$ except for the normalising constant, and there is similar agreement for $\bar{x}$ near -1. Actually $\log_e g_n(\bar{x})$ is remarkably close to $\log_e f_n(\bar{x})$ for quite moderate values of n over the whole range of x. Table 1 shows the agreement for $n = 6$, which could be improved by adjusting the normalising constant. With n as low as 6, $g_n(\bar{x})$ never differs from $f_n(\bar{x})$ by as much as 4 per cent. This example leads one to

TABLE 1

$\bar{x}$	.1	.2	.3	.4	.5	.6	.7	.8	.9
$\log_e f_6(\bar{x})$	0.419	0.172	−0.249	−0.860	−1.687	−2.778	−4.216	−6.243	−9.709
$\log_e g_6(\bar{x})$	0.445	0.199	−0.221	−0.829	−1.653	−2.742	−4.188	−6.228	−9.695
Difference	0.026	0.027	0.028	0.031	0.034	0.036	0.028	0.015	0.014

enquire under what conditions $f_n(\bar{x})/g_n(\bar{x}) \to 1$ *uniformly for all* $\bar{x}$ as $n \to \infty$, so that the relative accuracy of the approximation is maintained up to the ends of the range of $\bar{x}$. In Section 7 we show that the result is true for a wide class of density functions.

6. The real roots of $K'(T) = \xi$. In this section we discuss the existence and properties of the real roots of the equation $K'(T) = \xi$, upon which the approximation $g_n(\bar{x})$ depends. The conditions are here relaxed so that the distribution need not have a density function. The moment generating function is still assumed to satisfy the conditions of Section 2, namely that

$$M(T) = e^{K(T)} = \int_{-\infty}^{\infty} e^{Tx}\, dF(x)$$

converges for real T in $-c_1 < T < c_2$ where $0 \leqq c_1 \leqq \infty$ and $0 \leqq c_2 \leqq \infty$ but $c_1 + c_2 > 0$. Throughout this section T is supposed to take real values only.

The distribution may extend from $-\infty$ to ∞, or it may be limited at either or both ends. We shall write

$$
\begin{aligned}
F(x) &= 0, && x < a, \\
0 < F(x) &< 1, && a < x < b, \\
F(x) &= 1, && b < x,
\end{aligned}
$$

where if desired $a = -\infty$ or $b = \infty$, or both. Note that $b < \infty$ implies $c_2 = \infty$ so that $c_2 < \infty$ implies $b = \infty$, and similarly for a and c_1. The converse is not true since b and c_2 (or a and c_1) can both be infinite.

We now establish the conditions under which $K'(T) = \xi$ has no real root when ξ lies outside the interval (a, b), and has a unique simple root T_0 for every ξ in (a, b). It is convenient to consider first the case where both a and b are finite.

THEOREM 6.1. *$F(x) = 0$ for $x < a$, and $F(x) = 1$ for $x > b$ if and only if $K(T)$ exists for all real T and $K'(T) = \xi$ has no real root whenever $\xi < a$ or $\xi > b$.*

PROOF. Write

$$M(T, \xi) = e^{K(T)-T\xi} = \int_{-\infty}^{\infty} e^{T(x-\xi)}\, dF(x).$$

If $dF(x) = 0$ outside (a, b) then $M(T, \xi)$ exists for all real T and

$$M'(T, \xi) = \int_a^b (x - \xi)e^{T(x-\xi)}\, dF(x)$$

exists and has constant sign for all T when $\xi < a$ or $\xi > b$, and $K'(T) = \xi$ has then no real root.

Conversely, suppose $K(T)$ exists for all T and $K'(T) = \xi$ has no real root when $\xi < a$ or $\xi > b$. Then $M'(T, \xi)$ has constant sign in the domains $\xi < a$, $-\infty < T < \infty$ and $\xi > b$, $-\infty < T < \infty$ so that $M(T, \xi)$ is monotonic in T for these values of ξ.

Moreover $M(T, \xi)$ must increase with T for all $\xi < a$, and decrease with T for all $\xi > b$. For if $M(T, \xi)$ increases with T, then $dF(x) = 0$ for every $x < \xi$, otherwise $M(-\infty, \xi) = \infty$ and if this were true for all $\xi > b$ we should have $dF(x) = 0$ for all x. Similarly $M(T, \xi)$ cannot decrease with T for $\xi < a$.

Hence when $\xi < a$, $dF(x) = 0$ for all $x < \xi$, that is $F(x) = 0$ for all $x < a$. In the same way $F(x) = 1$ for all $x > b$.

THEOREM 6.2. *Let $F(x) = 0$ for $x < a$, $0 < F(x) < 1$ for $a < x < b$, $F(x) = 1$ for $b < x$, where $-\infty < a < b < \infty$. Then for every ξ in $a < \xi < b$ there is a unique simple root T_0 of $K'(T) = \xi$. As T increases from $-\infty$ to ∞, $K'(T)$ increases continuously from $\xi = a$ to $\xi = b$.*

PROOF. When $a < \xi < b$, $M'(-\infty, \xi) = -\infty$ and $M'(\infty, \xi) = \infty$, and $M'(T, \xi)$ is strictly increasing with T since $M''(T) > 0$. So for each ξ in $a < \xi < b$ there is a unique root T_0 of $M'(T, \xi) = 0$ and hence of $K'(T) = \xi$. Also $K''(T_0) = M''(T_0, \xi)/M(T_0, \xi)$ so that $0 < K''(T_0) < \infty$, and T_0 is a simple root and $K'(T_0)$ is a strictly increasing function of T_0.

For all T, $M'(T, b) < 0$ and so $K'(T) < b$, but $M'(T, b - \epsilon) \to \infty$ as $T \to \infty$

for every $\epsilon > 0$ so that $K'(T) > b - \epsilon$ for all sufficiently large T. Hence $K'(T) \to b$ as $T \to \infty$. Similarly $K'(T) \to a$ as $T \to -\infty$. This also implies $K''(T) \to 0$ as $T \to \pm\infty$.

The theorem has an obvious interpretation in terms of the family of conjugate distributions (the term is due to Khinchin [14])

$$dF(x, T) = Ce^{Tx}dF(x)$$

which have mean $K'(T)$ and variance $K''(T)$.

A complication arises when a and b are allowed to be infinite. Suppose for example that a is finite but $b = \infty$, so that $K(T)$ exists in $-\infty < T < c_2$ where $0 \leqq c_2 \leqq \infty$. If $c_2 = \infty$, then $K'(T) \to \infty$ as $T \to \infty$ and the theorems still hold, for however large ξ is, $M'(T, \xi) \to \infty$ as $T \to \infty$ and so $K'(T) > \xi$ for all sufficiently large T.

But if $c_2 < \infty$ the corresponding theorems do not hold without a further condition, for it is not necessarily true that $K'(T) \to \infty$ as $T \to c_2$. Consider the class of distributions

$$dF(x) = e^{-c_2 x}\, dG(x)$$

where $\int_a^\infty dG(x) = m_0 < \infty$ and $\int_a^\infty x dG(x) = m_1 < \infty$, but $\int_a^\infty e^{\epsilon x}\, dG(x) = \infty$,

for all $\epsilon > 0$. Here $K'(T)$ increases from $-\infty$ to m_1/m_0 as T increases from $-\infty$ to c_2, but $K'(T) = \infty$ for all $T > c_2$. So for $\xi > m_1/m_0$, $K'(T) = \xi$ has no real root though the distribution may extend to ∞.

The case $a = -\infty$ can be discussed similarly. In the general case where $K(T)$ exists in $-c_1 < T < c_2$ and a and b may be infinite, the conditions

$$\lim_{T \to -c_2} K'(T) = b, \qquad \lim_{T \to -c_1} K'(T) = a \tag{6.1}$$

are required for every ξ in (a, b) to have a corresponding T_0 in $(-c_1, c_2)$. They will be automatically satisfied except when a or b is infinite and the corresponding c_1 or c_2 is finite, in which case the appropriate condition has to be stated explicitly. But even when (6.1) is not satisfied the approximation $g_n(\bar{x})$ and the expansion (2.6) can still be used whenever $\bar{x}$ lies within the restricted range of values assumed by $K'(T)$.

7. Accuracy at the ends of the range of $\bar{x}$. We return to the distributions having a density function, and examine the accuracy of $g_n(\bar{x})$ and the expansion (2.6) for values of $\bar{x}$ near the ends of its admissible range (a, b), where the approximation might be expected to fail. It is assumed that the appropriate conditions hold for $K'(T) = \bar{x}$ to have a unique real root T_0 for every $\bar{x}$ in (a, b).

It has been proved that

$$| f_n(\bar{x})/g_n(\bar{x}) - 1 | < A(\bar{x})/n, \tag{7.1}$$

where $A(\bar{x})$ may depend on $\bar{x}$ since it is a function of T_0. The family of expansions (4.3) provides similar inequalities, and in particular an inequality of type (7.1) holds for symmetrical distributions when $g_n(\bar{x})$ is replaced by the limiting normal approximation to $f_n(\bar{x})$. But it is well known that the relative accuracy of the normal approximation, and of the Edgeworth series generally, deteriorates in most cases as $\bar{x}$ approaches the ends of its range. For example, if the interval (a, b) is finite and $f_n(\bar{x}) \to 0$ as $\bar{x} \to a$ or b, what corresponds to $A(\bar{x})$ in (7.1) becomes intolerably large as x approaches a or b, since the normal approximation can never be zero.

We now show that for a wide class of distributions $g_n(\bar{x})$ satisfies (7.1) with $A(\bar{x}) = A$, independent of $\bar{x}$, as $\bar{x}$ approaches a or b. In fact, for such distributions the asymptotic expansion of $f_n(\bar{x})/g_n(\bar{x})$ given by (2.6) is valid uniformly as $\bar{x} \to a$ or b. This will be so if $\lambda_j(T)$ remains bounded as $T \to -c_1$ or c_2 for every fixed j, so we examine the behaviour of $\lambda_j(T)$ near the ends of the interval. Equivalently, we study the conjugate distributions with density function

$$f(x, T) = Ce^{Tx}f(x) \tag{7.2}$$

whose jth cumulant is $K^{(j)}(T)$. The form of $f(x, T)$ as T approaches $-c_1$ or c_2 depends on the behaviour of $f(x)$ as x approaches a or b. For the commonest end conditions on $f(x)$, it will appear that $f(x, T)$ approximates either to the gamma form of Example 5.2 or to the normal form as $T \to -c_1$ or c_2. In the first case $\lambda_j(T)$ is bounded for given j; in the second case $\lambda_j(T) \to 0$ so that $g_n(\bar{x})$, for any n, becomes progressively more accurate as $\bar{x} \to b$, its relative error tending to a limiting value which is of smaller order than any power of n^{-1}.

We begin by discussing distributions with $b = \infty$ and first consider asymptotic forms of $f(x)$ when x is large for which $f(x, T)$ approximates to the gamma form.

EXAMPLE 7.1. $\qquad f(x) \sim Ax^{\alpha-1}e^{-cx}, \qquad \alpha > 0, c > 0.$

Let X be large. Then

$$M^{(j)}(T) = \int_{-\infty}^{\infty} x^j e^{Tx} f(x)\, dx = I_1 + I_2$$

where $I_1 = \int_{-\infty}^{X} x^j e^{Tx} f(x)\, dx$ is bounded as $T \to c$, and for small $c - T$,

$$I_2 \sim \int_X^{\infty} x^{j+\alpha-1} e^{-(c-T)x}\, dx = \frac{A}{(c-T)^{j+\alpha}} \int_{X(c-T)}^{\infty} w^{j+\alpha-1} e^{-w}\, dw$$

$$\sim A\,\Gamma(j+\alpha)/(c-T)^{j+\alpha}.$$

Thus

$$K^{(j)}(T) \sim \frac{\alpha}{(c-T)^j}; \qquad \lambda_j(T) \sim \alpha^{1-j/2} \tag{7.3}$$

for every j. In this case $f(x, T)$ tends to the gamma form as $T \to c$. The result is in fact a familiar Abelian theorem for Laplace transforms, and a more general form of it (Doetsch [7] p. 460) can be restated for our purpose as follows.

THEOREM 7.1. *Let* $f(x) \sim Ax^{\alpha-1}l(x)e^{-cx}$ *for* $\alpha > 0$ *and* $c > 0$, *where* $l(x)$ *is continuous and* $l(kx)/l(x) \to 1$ *as* $x \to \infty$ *for every* $k > 0$. *Then, as* $T \to c$,

$$M^{(j)}(T) \sim A\,\frac{\Gamma(j+\alpha)}{(c-T)^{j+\alpha}}\,l\left(\frac{1}{c-T}\right) \quad \text{and} \quad \lambda_j(T) \sim \alpha^{1-j/2}.$$

This enables us to include end conditions of the form $Ax^{\alpha-1}\log x\, e^{-cx}$ or $Ax^{\alpha-1}$ $\log\log x\, e^{-cx}$, etc. In all such cases $f(x, T)$ tends to the gamma form as $T \to c$.

In the second class of end conditions $f(x, T)$ approximates to the normal form for limiting values of T. We first consider heuristically some typical examples, again with $b = \infty$

EXAMPLE 7.2. $\quad f(x) \sim A \exp(\beta x^\alpha - cx), \qquad \beta > 0, c > 0, 0 < x < 1.$

Here we might expect $\lambda_j(T) \to 0$ as $T \to c$, for when $c - T$ is small the dominant part of $f(x, T)$ lies in the region of large x where

$$f(x, T) \sim CA \exp(\beta x^\alpha - (c - T)x).$$

This has a unique maximum at $x_0 = [\alpha\beta/(c - T)]^{1/(1-\alpha)}$ which is large for small $c - T$. If we put $x = x_0 y$ the corresponding density for y is $c' \exp[\beta x_0^\alpha (y^\alpha - \alpha y)]$ which has a sharp maximum at $y = 1$, near which it approximates to the normal form $c'' \exp[-\frac{1}{2}\beta\alpha(1 - \alpha)x_0^\alpha(y - 1)^2]$; it is relatively negligible elsewhere.

EXAMPLE 7.3. $\quad f(x) \sim A \exp(-\beta x^\alpha), \qquad \beta > 0, \alpha > 1.$

In this case T can be indefinitely large. We again expect $\lambda_j(T) \to 0$ as $T \to \infty$, for

$$f(x, T) \sim CA \exp[-\beta x^\alpha + Tx]$$

has a unique maximum at $x_0 = (T/\alpha\beta)^{1/(\alpha-1)}$ which tends to infinity with T; with $x = z_0 y$ the density for y becomes $c' \exp[\beta x_0^\alpha(y^\alpha - \alpha y)]$, which approximates to $c'' \exp[-\frac{1}{2}\beta\alpha(\alpha - 1)x_0^\alpha(y - 1)^2]$ as before.

These examples are included in the following general theorem concerning end conditions of the type $f(x) \sim e^{-h(x)}$, where $x^2h''(x) \to \infty$ as $x \to \infty$. Subject to a restriction on the variation of $h''(x)$ it is shown that $\lambda_j(T) \to 0$ in such cases as T tends to its upper limit.

THEOREM 7.2. *Let* $f(x) \sim e^{-h(x)}$ *for large* x, *where* $h(x) > 0$ *and* $0 < h''(x) < \infty$. *Let* $v(x)$ *and* $w(x)$ *exist such that*

$$\text{(i)} \quad [v(x)]^2h''(x) \to \infty \qquad \text{(ii)} \quad e^{-w(x)}h''(x) \to 0$$

monotonically as $x \to \infty$, *where*

$$v(x) > 0, \qquad |v'(x)| \leqq \alpha < \infty, \qquad w(x) = \int (1/v(x))\,dx.$$

Then $\lambda_j(T) \to 0$ *as* T *tends to its upper limiting value.*

Examples 7.2 and 7.3 are covered by $v(x) = x/\gamma$ for some $\gamma > 0$, conditions (i) and (ii) reducing to $x^2h''(x) \to \infty$, and $x^{-\gamma}h''(x) \to 0$. For $h(x) = e^x$ one can take $v(x) = \frac{1}{2}$, for $h(x) = e^{e^x}$ take $v(x) = \frac{1}{2}e^{-x}$, and so on. In all cases $v(x)/x$ is bounded and $w(x)$ increases at least as fast as log x.

Since $0 < h''(x) < \infty$, $h'(x)$ is strictly increasing and $h'(x) \to c \leqq \infty$ as $x \to \infty$. Thus for large x, $f(x, T) \sim Ce^{Tx-h(x)}$ has a single maximum at the unique root x_0 of $h'(x_0) = T$, where $x_0 \to \infty$ as $T \to c \leqq \infty$.

The jth moment of $f(x, T)$ about x_0 is

$$\mu_j(T) = C \int_{-\infty}^{\infty} (x - x_0)^j f(x) e^{Tx}\, dx.$$

It will be shown that as $T \to c$ the major contribution to the integral comes from within a range $x_0 \pm \epsilon v(x_0)$ where ϵ is arbitrarily small. Consider first the behaviour of $v(x)$ and $w(x)$ in this interval. Since $| v'(x) | < \alpha$ as $x \to \infty$ we have for large x_0 and $| x - x_0 | \leqq \epsilon v(x_0)$,

$$| v(x) - v(x_0) | \leqq \alpha | x - x_0 | < \alpha\epsilon v(x_0)$$

that is

$$| v(x)/v(x_0) - 1 | < \alpha\epsilon. \tag{7.4}$$

Also for some x_1 in (x, x_0),

$$w(x) - w(x_0) = (x - x_0)w'(x_1) = (x - x_0)/v(x_1)$$

so that for $| x - x_0 | < \epsilon v(x_0)$,

$$| w(x) - w(x_0) | \leqq \epsilon \frac{v(x_0)}{v(x_1)} \leqq \frac{\epsilon}{1 - \alpha\epsilon}. \tag{7.5}$$

Let X be large, but choose T so that $x_0 > X + j/T$. Then

$$\mu_j(T) \sim C \int_{-\infty}^{X} (x - x_0)^j f(x) e^{Tx}\, dx$$
$$+ C\left\{\int_X^{x_0-\epsilon v(x_0)} + \int_{x_0-\epsilon v(x_0)}^{x_0+\epsilon v(x_0)} + \int_{x_0+\epsilon v(x_0)}^{\infty}\right\} [(x - x_0)^j \exp [xh'(x_0) - h(x)]\, dx$$
$$= I_1 + I_2 + I_3 + I_4$$

say. We examine the magnitude of each term as $T \to c$.

Since $(x_0 - x)^j e^{Tx}$ has its maximum at $(x_0 - j/T) > X$,

$$| I_1 | < C(x_0 - x)^j e^{TX} F(X) < Cx_0^j e^{Xh'(x_0)}$$

For I_2,

$$| I_2 | = Ce^{x_0h'(x_0)-h(x_0)} \int_X^{x_0-\epsilon v(x_0)} (x_0 - x)^j e^{-\psi(x,x_0)}\, dx \tag{7.6}$$

where we write $\psi(x, x_0) = h(x) - h(x_0) - (x - x_0)h'(x_0)$. By condition (ii) of the theorem, when $x \leqq x_0$,

$$h''(x) \geqq e^{w(x)-w(x_0)}h''(x_0) > 0$$

and for $x \leqq x_0 - \epsilon v(x_0)$,

$$h'(x_0) - h'(x) > h''(x_0) \int_{x_0-\epsilon v(x_0)}^{x_0} e^{w(x)-w(x_0)}\, dx$$
$$\geqq \eta v(x_0)h''(x_0),$$

where $\eta = \epsilon e^{-\epsilon/(1-\alpha\epsilon)}$, by (7.5). So $\psi(x, x_0) > \eta v(x_0)h''(x_0)(x_0 - x)$, and from (7.6),

$$| I_2 | < Ce^{x_0 h'(x_0)-h(x_0)}/[\eta v(x_0)h''(x_0)]^{j+1}.$$

For I_3,

$$I_3 = Ce^{x_0 h'(x_0)-h(x_0)}\left\{\int_{x_0-\epsilon v(x_0)}^{x_0} + \int_{x_0}^{x_0+\epsilon v(x_0)}\right\} (x - x_0)^j e^{-\psi((x,x_0)} \, dx$$
$$= Ce^{x_0 h'(x_0)-h(x_0)}\{J_1 + J_2\},$$

say. When $x_0 - \epsilon v(x_0) \leqq x \leqq x_0$ we have from (i) and (ii),

$$e^{w(x)-w(x_0)} \leqq h''(x)/h''(x_0) \leqq [v(x_0)/v(x)]^2 \tag{7.7}$$

and so from (7.4) and (7.5),

$$\tfrac{1}{2}h''(x_0)(x - x_0)^2 e^{-\epsilon/(1-\alpha\epsilon)} \leqq \psi(x, x_0) \leqq \tfrac{1}{2}h''(x_0)(x - x_0)^2(1 + \alpha\epsilon)^2$$

Putting $u = (x - x_0)[h''(x_0)]^{1/2}$ in J_1 makes the lower limit of integration become $-\epsilon v(x_0)[h''(x_0)]^{1/2}$, which tends to $-\infty$ as $x_0 \to \infty$ for fixed ϵ, by (i). Hence

$$J_1 \sim (-)^j \frac{2^{(j-1)/2}\Gamma[(j + 1)/2]}{[h''(x_0)]^{(j+1)/2}} \{1 + O(\epsilon)\}.$$

In the range $x_0 \leqq x \leqq x_0 + \epsilon v(x_0)$ the inequalities (7.7) are reversed and

$$\tfrac{1}{2}h''(x_0)(x - x_0)^2(1 - \alpha\epsilon)^2 \leqq \psi(x, x_0) \leqq \tfrac{1}{2}h''(x_0)(x - x_0)^2 e^{\epsilon/(1-\alpha\epsilon)}$$

so that

$$J_2 \sim \frac{2^{(j-1)/2}\Gamma[(j + 1)/2]}{[h''(x_0)]^{(j+1)/2}} \{1 + O(\epsilon)\}.$$

Hence if j is even

$$I_3 \sim C \frac{e^{x_0 h'(x_0)-h(x_0)}}{[h''(x_0)]^{(j+1)/2}} \frac{(2j)!}{2^j j!} (2\pi)^{1/2}\{1 + O(\epsilon)\}$$

while if j is odd

$$I_3 \sim C \frac{e^{x_0 h'(x_0)-h(x_0)}}{[h''(x_0)]^{(j+1)/2}} \cdot O(\epsilon).$$

For I_4,

$$I_4 = e^{x_0 h'(x_0) - h(x_0)} \int_{x_0+\epsilon v(x_0)}^{\infty} (x - x_0)^j e^{-\psi(x, x_0)}\, dx.$$

The inequality $h''(x) \geqq [v(x_0)/v(x)]^2 h''(x_0) > 0$ shows, as with I_2, that

$$I_4 < \frac{Ce^{x_0 h'(x_0) - h(x_0)} j!}{[\epsilon(1 - \alpha\epsilon)v(x_0)h''(x_0)]^{j+1}}.$$

We now show that I_3 is the dominant term. First let j be even. As $T \to c$, both I_2/I_3 and I_4/I_3 are $O\{[v^2(x_0)h''(x_0)]^{-(j+1)/2}\}$, and so $\to 0$ for fixed ϵ. Further,

$$| I_1 |/I_3 < x_0^j [h''(x_0)]^{(j+1)/2} e^{h(X) - \psi(X, x_0)}.$$

From (ii), $h''(x_0) < e^{w(x_0)}$ as x_0 becomes large. Also since $v(x)/x$ is bounded, (i) implies that $(x - X)h''(x)v(x) \to \infty$ and so for all large enough x_0,

$$\psi(X, x_0) = \int_X^{x_0} (x - X)h''(x)\, dx > A \int_X^{x_0} (1/v(x))\, dx = A\{w(x_0) - w(X)\}$$

whatever $A > 0$. Thus

$$| I_1 |/I_3 = O\{\exp[j \log x_0 - [A - \tfrac{1}{2}(j + 1)]w(x_0)]\}$$

which tends to zero as $T \to c$ for fixed X if A is large enough, since $w(x)$ increases at least as fast as $\log x$.

It follows that for even j,

$$\mu_j(T) \sim \frac{Ce^{x_0 h'(x_0) - h(x_0)}}{[h''(x_0)]^{(j+1)/2}} \cdot \frac{(2j)!}{2^j j!} (2\pi)^{1/2}$$

$$\sim [h''(x_0)]^{-j/2} \frac{(2j)!}{2^j j!}$$

since $\mu_0(T) = 1$. Similarly when j is odd,

$$\mu_j(T) \sim [h''(x_0)]^{-j/2}\, O(\epsilon)$$

as $T \to c$, so the odd moments can be made relatively negligible for arbitrarily small ϵ. Thus the moments tend to those of the normal distribution and $\lambda_j(T) \to 0$ as $T \to c$.

Turning now to the case where $x \leqq b < \infty$ we consider forms of $f(x)$ when $b - x$ is small. Again there are found to be two classes of end conditions for which $\lambda_j(T)$ is bounded as $T \to \infty$, where $f(x, T)$ tends respectively to the gamma and to the normal form. It is convenient to put $u = b - x$ and regard $(-)^j K^{(j)}(T)$ as the jth cumulant of the distribution of u with density $f(b - u, T) = Be^{-Tu} f(b - u)$ for $u \geqq 0$.

EXAMPLE 7.4. $\quad f(x) \sim A(b - x)^{\alpha-1}, \quad \alpha > 0.$

The jth moment of u about its origin is

$$B \int_0^\infty u^j e^{-Tu} f(b-u)\,du \sim BA \int_0^\delta u^{j+\alpha-1} e^{-Tu}\,du + B \int_\delta^\infty u^j e^{-Tu} f(b-u)\,du$$

$$\sim BA \frac{\Gamma(\alpha+j)}{T^{\alpha+j}} \qquad T \to \infty,$$

where δ is small, the remainder being $O(e^{-T\delta})$ for large T. It follows that $\lambda_j(T) \sim (-)^j \alpha^{1-j/2}$. As in Example 7.1 this is a well known result on Laplace transforms, and its more general form (Doetsch [7], p. 476) yields the following theorem.

THEOREM 7.3. *Let* $f(x) \sim A(b-x)^{\alpha-1} l(b-x)$ *for* $\alpha > 0$, *where* $l(u)$ *is continuous and* $[l(ku)]/l(u) \to 1$ *as* $u \to 0$ *for every* $k > 0$. *Then* $\lambda_j(T) \sim (-)^j \alpha^{1-j/2}$.

For example $l(u)$ could be $\log(1/u)$ or $\log\log(1/u)$, etc.

The second class of end conditions is typified by the following example.

EXAMPLE 7.5. $\qquad f(x) \sim A \exp[-\beta/(b-x)^\gamma], \qquad \beta > 0, \gamma > 0.$

As in Example 7.2 we expect $\lambda_j(T) \to 0$ as $T \to \infty$, for

$$Ce^{-Tu} f(b-u) \sim CA \exp[-Tu - \beta/u^\gamma]$$

has a unique maximum at $u_0 = (\beta\gamma/T)^{1/(\gamma+1)}$, and the density function for $y = u/u_0$ is

$$C' \exp[-\beta u_0^{-\gamma}(\gamma y + y^{-\gamma})] \sim C'' \exp[-\tfrac{1}{2}\beta\gamma(\gamma+1)u_0^{-\gamma}(y-1)^2]$$

The general theorem analogous to Theorem 7.2 is:

THEOREM 7.4. *Let* $f(x) \sim e^{-h(x)}$ *for small* $b - x$, *where* $h(x) > 0$ *and* $0 < h''(x) < \infty$. *Let* $v(u)$ and $w(u)$ *exist such that*

$$\text{(i) } [v(b-x)]^2 h''(x) \to \infty, \qquad \text{(ii) } e^{w(b-x)} h''(x) \to 0,$$

monotonically as $x \to b$, *where* $v(0) = 0$ *and* $w(u) = \int [1/v(u)]\,du$, *and* $0 < v'(u) \leqq \alpha < \infty$ *for* $u > 0$. *Then* $\lambda_j(T) \to 0$ *as* $T \to \infty$.

As before $h'(x)$ is strictly increasing, and $h'(x) \to \infty$ as $x \to b$ since (i) implies $(b-x)^2 h''(x) \to \infty$, and $h'(x_0) = T$ has a unique root x_0 where $x_0 \to b$ as $T \to \infty$. Thus $f(b-u, T)$ has a unique maximum at $u_0 = b - x_0$ for large T, and $u_0 \to 0$ as $T \to \infty$. The jth moment of u about u_0 is

$$\mu_j(T) = B \int_0^\infty (u-u_0)^j e^{-Tu} f(b-u)\,du.$$

We write

$$\int_0^\infty = \int_0^{u_0 - \epsilon v(u_0)} + \int_{u_0-\epsilon v(u_0)}^{u_0+\epsilon v(u_0)} + \int_{u_0+\epsilon v(u_0)}^{\delta} + \int_\delta^\infty$$

where ϵ and δ are small. The proof then proceeds with appropriate modifications as in Theorem 7.2.

8. Discrete variables. The discussion has so far been concerned with approximations to probability densities, but the saddlepoint method provides similar approximations to probabilities when the variable is discrete, and indeed it is typically used for this purpose in statistical mechanics. Consider, for example, a variable x which takes only integral values $x = r$ with nonzero probabilities $p(r)$. The moment generating function,

$$M(T) = e^{K(T)} = \sum_r p(r)e^{Tr} \tag{8.1}$$

is assumed to satisfy conditions (6.1).

The mean $\bar{x}$ of n independent x's can take only values $\bar{x} = r/n$, for which the probabilities are

$$p_n(\bar{x}) = \frac{1}{2\pi i}\int_{\tau-i\pi}^{\tau'+i\pi} e^{n[K(T)-T\bar{x}]}\, dT \tag{8.2}$$

analogous to (2.2). The contour is again chosen to be the line $T = T_0 + iy$ passing through the unique real saddle point T_0, but it now terminates at $T_0 \pm i\pi$. This ensures that the integrand attains its maximum modulus at T_0 but nowhere else on the contour, provided we exclude cases where $p(r) = 0$ except at multiples of an integer greater than unity. The discussion of Section 2 shows that the maximum modulus is attained when y satisfies $\cos(ry - \alpha) = 1$ for some α and all integral r, and $y = 0$ is the only possible value in $(-\pi, \pi)$. The argument then proceeds as before and leads to the approximation

$$p_n(\bar{x}) \sim \frac{e^{n[K(T_0)-T_0\bar{x}]}}{[2\pi n K''(T_0)]^{1/2}}\{1 + O(n^{-1})\} \tag{8.3}$$

where $\bar{x} = r/n$ and r is an integer.

As an example, consider the binomial distribution

$$p(r) = \binom{N}{r}(1-p)^r p^{n-r}.$$

Here

$$K(T) = N\log\{1 + p(e^T - 1)\}, \qquad K'(T_0) = Npe^{T_0}/[1 + p(e^{T_0} - 1)] = \bar{x},$$

$$e^{T_0} = [\bar{x}/(N-\bar{x})]\cdot[(1-p)/p], \quad K''(T_0) = \bar{x}(N-\bar{x})/N,$$

$$p_n(\bar{x}) \sim \frac{N^{nN+1/2}}{(2\pi n)^{1/2}}\frac{(1-p)^{n(N-\bar{x})}p^{n\bar{x}}}{(N-\bar{x})^{n(N-\bar{x})+1/2}\bar{x}^{n\bar{x}+1/2}}\{1 + O(n^{-1})\}.$$

This is the familiar intermediate form obtained on replacing the factorials by Stirling's approximation before passing to the normal limit.

9. Ratio of sums of random variables. The saddlepoint technique can also be applied to the distribution of ratios. Cramér (4) has shown that if x and y are two independent random variables with densities $f_1(x)$ and $f_2(y)$ and characteristic functions $\phi_1(t)$ and $\phi_2(u)$, and if $y \geqq 0$, the density function for $r = x/y$ is given by

$$f(r) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \phi_1(t)\phi_2'(-rt)\, dt \tag{9.1}$$

provided y has a finite mean. (Gurland [11] relaxes this condition by introducing principal values. Cramér states the condition differently and appears to require unnecessarily that x shall have a finite mean also.) Cramér deduced the result from the distribution of $x - ry$ for fixed r, but it also follows on applying Parseval's theorem to the formula

$$f(r) = \int_0^{\infty} f_1(ry)f_2(y)y\, dy \tag{9.2}$$

where y must have a finite mean to make $\phi_2'(-rt)$ the Fourier transform of $iyf_2(y)$. In terms of cumulant generating functions (9.1) takes the form

$$f(r) = \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau'+i\infty} e^{K_1(T)+K_2(-rT)}\, K_2'(-rT)\, dT.$$

Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ be independent random samples from these distributions, their sums being X and Y. The density for $R = x/y$ is then

$$f_{n_1,n_2}(R) = \frac{n_2}{2\pi i} \int_{\tau-i\infty}^{\tau'+i\infty} e^{n_1K_1(T)+n_2K_2(-RT)}\, K_2'(-RT)\, dT.$$

When n_1 and n_2 are large, an approximation is found by passing the path of integration through a saddlepoint T_0 of the exponential part of the integrand, given by

$$n_1K_1'(T_0) - n_2RK_2'(-RT) = 0 \tag{9.3}$$

Assuming conditions (6.1) to be satisfied, both $K_1'(T)$ and $K_2'(T)$ are increasing functions of T taking every admissible value of X and Y respectively as T varies over its appropriate interval, so that to every R there is a single real root T_0 of (9.3). (However, it is possible for the same T_0 to correspond to more than one value of R, since $TK_2'(T)$ is not necessarily monotonic and so dT_0/dR may change sign). Proceeding as before, expanding $K_2'(-RT)$ also, we obtain an asymptotic expansion whose dominant term is

$$g_{n_1,n_2}(R) = \frac{n_2K_2'(-RT_0)e^{n_1K_1(T_0)+n_2K_2(-RT_0)}}{\{2\pi[n_1K_1''(T_0) + n_2R^2K_2''(-RT_0)]\}^{1/2}}$$

the remainder being relatively $O(n^{-1})$ where $n = \min(n_1, n_2)$.

EXAMPLE 9.1. $f_1(x) = A_1x^{\alpha_1-1}e^{-\beta_1x}$, $\quad f_2(y) = A_2y^{\alpha_2-1}e^{-\beta_2y}$,

where x, y, α_1, β_1, α_2, β_2 are all positive. In this case

$$T_0 = \frac{1}{R}\frac{(n_2\alpha_2\beta_1R - n_1\alpha_1\beta_2)}{(n_1\alpha_1 + n_2\alpha_2)}.$$

The approximation is found to be

$$g_{n_1,n_2}(R) = \frac{\beta_1^{n_1\alpha_1}\beta_2^{n_2\alpha_2}(n_1\alpha_1 + n_2\alpha_2)^{n_1\alpha_1+n_2\alpha_2-1/2}}{(2\pi)^{1/2}(n_1\alpha_1)^{n_1\alpha_1-1/2}(n_2\alpha_2)^{n_2\alpha_2-1/2}} \frac{R^{n_1\alpha_1-1}}{(\beta_1 R + \beta_2)^{n_1\alpha_1+n_2\alpha_2}},$$

which differs from the exact density function only in the normalising constant, and so is "exact" in the sense of Example 5.2. This suggests that there may again be a class of distributions for which the relative error is bounded uniformly over the whole range of R for every n.

An extension of (9.1) is available when the variables are not independent (Cramér [6] p. 317, ex. 6; Geary [10]). If (x, y) has a bivariate density function $f(x, y)$ everywhere and characteristic function $\phi(t, u)$, and if $y \geqq 0$, the density function for $r = x/y$ is

$$f(r) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{\partial\phi(t, u)}{\partial u}\right]_{u=-rt} dt \tag{9.4}$$

provided the integral is absolutely convergent, which requires y to have a finite mean. The following proof of (9.4) shows the integrand to be proportional to a characteristic function which attains its maximum modulus only at $t = 0$, so that the previous methods are applicable. Corresponding to (9.2) we have

$$f(r) = \int_0^{\infty} f(ry, y)y\, dy. \tag{9.5}$$

Write $\eta = E(y)$ and define a new distribution with density and characteristic function

$$h(x, y) = \frac{1}{\eta} yf(x, y) \qquad \phi(x, y) = \frac{1}{\eta}\frac{\partial\phi(t, u)}{i\partial u}. \tag{9.6}$$

From (9.5) it is seen that $(1/\eta)f(r)$ can be regarded as the probability density at zero of the variable $w = x - ry$, where (x, y) has the distribution (9.6). The result then follows from the fact that w has the characteristic function

$$\frac{1}{\eta}\left[\frac{\partial\phi(t, u)}{i\partial u}\right]_{u=-rt}.$$

For a random sample of n, the ratio R of the sums X and Y has density

$$f_n(R) = \frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau'+i\infty} e^{nK(T,-RT)}\left[-\frac{1}{T}\frac{\partial K(T, -RT)}{\partial R}\right]\partial T$$

in terms of the bivariate cumulant generating function. The saddlepoint approximation is

$$g_n(R) = \left\{\frac{n}{2\pi K''(T_0, -RT_0)}\right\}^{1/2} e^{nK(T_0,-RT_0)}\left[\frac{-1}{T_0}\frac{\partial K(T_0, -RT_0)}{\partial R}\right]$$

where

$$\frac{\partial K(T_0, -RT_0)}{\partial T_0} = 0.$$

EXAMPLE 9.2. Let $x = \frac{1}{2}u^2$ and $y = \frac{1}{2}v^2$, where u and v have a bivariate normal distribution with unit variances and correlation coefficient ρ. Thus $R = X/Y$ is a "variance ratio" calculated from two equal correlated samples. The exact distribution of R has been given by Bose [1] and Finney [8]. We find

$$K(T, -RT) = \tfrac{1}{2}\log\{1 + (R-1)T - RT^2(1-\rho^2)\},$$

$$T_0 = \frac{(R-1)}{2R(1-\rho^2)}, \qquad K''(T_0, -RT_0) = \frac{4R(1-\rho^2)^2}{[(1+R)^2 - 4\rho^2 R]},$$

$$\frac{-1}{T_0}\frac{\partial K(T_0, -RT_0)}{\partial R} = \frac{(R+1)(1-\rho^2)}{[(1+R)^2 - 4\rho^2 R]},$$

$$g_n(R) = 2^{n-1}\left(\frac{n}{2\pi}\right)^{1/2}\frac{(1-\rho^2)^{n/2}R^{(n/2)-1}(1+R)}{[(1+R)^2 - 4\rho^2 R]^{(n+1)/2}}$$

which again agrees with the exact distribution except for the normalising constant.

In the most general situation where the sample members are themselves correlated, the saddlepoint method can still be applied. In each particular case the contribution to the integral from parts of the contour outside a neighbourhood of the saddlepoint must be established as negligible. One can obtain, in this way an approximation to the distribution of the sample serial correlation coefficient of lag 1 from a linear Markoff population. With the usual "circular" definitions it turns out to be the approximation given by Leipnik [15], but a similar approximation can also be found for the noncircular case. A detailed account of this work will appear elsewhere.

10. Acknowledgments. I am much indebted to D. V. Lindley and Sir Harold Jeffreys for many stimulating discussions and useful suggestions, to a referee for comments which led to improvements in the paper, and to D. A. East for computing Table 1.

11. Appendix. The identity of the series (2.6) and 3.3) may be established as follows. For the contour $T = T_0 + iy$ the inversion formula is

$$f_n(\bar{x}) = \frac{n}{2\pi} e^{n[K(T_0) - T_0\bar{x}]} \int_{-\infty}^{\infty} e^{-nw^2/2}\, dy$$

where w^2 is defined by (3.2). With $v = y[nK''(T_0)]^{1/2}$ and $s = n^{-1/2}$ this becomes

$$f_n(\bar{x}) = \frac{1}{2\pi}\left[\frac{n}{K''(T_0)}\right]^{1/2} e^{n[K(T_0) - T_0\bar{x}]} \int_{-\infty}^{\infty} e^{-w^2(ivs)/2s^2}\, dv$$

with $z = ivs$ in (3.2). To get (2.5) the integrand is expanded as a power series in s. Term-by-term integration gives (2.6). Thus

$$\exp\left[-\frac{w^2(ivs)}{2s^2}\right] = \sum_{m=0}^{\infty} b_m(v)s^m$$

where

$$b_m(v) = \frac{1}{m!}\frac{\partial^m}{\partial s^m}\exp\left[-\frac{w^2(ivs)}{2s^2}\right]\Bigg|_{s=0}$$
$$= \frac{1}{m!}v^m\frac{\partial^m}{\partial x^m}\exp\left[-\frac{v^2w^2(ix)}{2x^2}\right]\Bigg|_{x=0}$$

Since $w^2(ix)/x^2 \sim 1 + O(x)$, for small x we can interchange the order of differentiation with respect to x and integration with respect to v. Only the even terms survive and

$$\int_{-\infty}^{\infty} b_{2r}(v)\,dv = \frac{1}{(2r)!}\frac{d^{2r}}{dx^{2r}}\int_{-\infty}^{\infty} v^{2r}\exp\left[-\frac{v^2w^2(ix)}{2x^2}\right]dv\,\Bigg|_{x=0}$$
$$= \frac{(2\pi)^{1/2}}{2^r r!}\frac{d^{2r}}{dx^{2r}}\left[\frac{x}{w(ix)}\right]^{2r+1}\Bigg|_{x=0} = (2\pi)^{1/2}a_r,$$

putting $z = ix$ in (3.4).

REFERENCES

[1] S. Bose, "On the distribution of the ratio of variance of two samples drawn from a given normal bivariate correlated population," *Sankhya*, Vol. 2 (1935), pp. 65–72.
[2] D. R. Cox, "A note on the asymptotic distribution of the range," *Biometrika*, Vol. 35 (1948), pp. 311–315.
[3] H. Cramér, "On the composition of elementary errors," *Skand. Aktuarietids.*, Vol. 11 (1928), pp. 13–74, 141–180.
[4] H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, 1937.
[5] H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilités," *Actualités Scientifiques et Industrielles*, No. 736, Hermann & Cie, Paris, 1938.
[6] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[7] G. Doetsch, *Handbuch der Laplace Transformation*, Band I, Birkhäuser, Basel, 1950.
[8] D. J. Finney, "The distribution of the ratio of estimates of two variances in a sample from a normal bivariate population," *Biometrika*, Vol. 30 (1938), pp. 190–192.
[9] R. H. Fowler, *Statistical Mechanics*, Cambridge University Press, 1936.
[10] R. C. Geary, "Extension of a theorem by Harald Cramér on the frequency distribution of a quotient of two variables," *J. Roy. Stat. Soc.*, Vol. 17 (1944), pp. 56–57.
[11] J. Gurland, "Inversion formulae for the distribution of ratios," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 228–237.
[12] H. Jeffreys, *Theory of Probability*, Oxford University Press, 1948.
[13] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, Cambridge University Press, 1950.
[14] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, Inc., New York, 1949.
[15] R. B. Leipnik, "Distribution of the serial correlation in a circularly correlated universe," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 80–87.
[16] H. L. Seal, "Spot the prior reference," *J. Inst. of Actuaries, Students' Soc.*, Vol. 10 (1951), pp. 255–258.
[17] G. N. Watson, *Theory of Bessel Functions*, Cambridge University Press, 1948.
[18] A. Wintner, *The Fourier Transforms of Probability Distributions*, Johns Hopkins University, Baltimore, 1947.