SADDLE-POINT METHODS FOR THE MULTINOMIAL DISTRIBUTION[1]

By I. J. Good

1. Summary. Many problems in the theory of probability and statistics can be solved by evaluating coefficients in generating function, or, for continuous differentiable distributions, by an analogous process with Laplace or Fourier transforms. As pointed out for example by H. E. Daniels [2], these problems can often be solved by asymptotic series derived by the saddle-point method from integrals containing a large parameter. Daniels gave a form of saddle-point theorem that is convenient for applications to probability and statistics. In the present paper we extend the theorem in various directions and give some applications to distributions connected with the multinomial distribution, especially to the distribution of χ^2 and to the distribution of the maximum entry in a multinomial distribution.

2. Introduction. The use of asymptotic formulae in practical statistics has, historically, been partly experimental. As much regard has been paid to numerical examples as to analytical bounds on the errors. Analytical bounds have a habit of being much larger than life, and are often of less practical value than the second term of an asymptotic expansion. If the second term is small then we can be happier about relying on the first term. If it is not small then we learn even more, and we also become interested in finding the third term. There is therefore a definite need in statistics for two-term and three-term asymptotic expansions.

In Sec. 6 we give three rather general theorems about asymptotic expansions of integrals, double integrals, and multidimensional integrals, in a form convenient for statistical applications. These theorems are adequately motivated by the earlier sections. In Sec. 3 we make some preliminary remarks about the multinomial distribution and tests for it. In Sec. 4 we give some examples to show how generating functions arise for these tests. In Sec. 5 we give brief descriptions of continuous and discrete methods of extracting coefficients from generating functions, and the continuous method is elaborated in Sec. 6, where the general theorems on asymptotic expansions are given. These theorems are applied in Secs. 7 and 8 to the distribution of the maximum entry and to that of χ^2 for a multinomial distribution. (When we refer to χ^2 we mean the statistic for testing goodness of fit, not the gamma-variate.) In Secs. 9 and 10 we give some examples of the discrete method of extracting coefficients and some combinatorial formulae for the distribution of the maximum entry.

Received December 10, 1956.

[1] This work was sponsored in part by the Office of Naval Research while the author was a Research Associate at Princeton University in 1955. I am indebted to Professor J. W. Tukey and to the Referee for making a number of useful suggestions when the paper was in draft.

3. Significance tests for multinomial distributions. Multinomial distributions arise, for example, in the following problems.

(a) Testing the goodness of fit of observations to a theoretical continuous distribution.

(b) Testing whether a sequence of non-negative integral variables arose from independent observations of a Poisson variable of unknown mean. (See, for example, Rao and Chakravarti [21] or Hoel [10], page 198.)

(c) Testing whether digits are adequate as random sampling numbers.

In all three of these applications the equiprobable multinomial distribution (with all cell probabilities equal) is of special interest and therefore, in the present paper, more attention will be given to this case than to the general multinomial distribution. Furthermore the theory is simpler for the special case.

Let our multinomial distribution have t categories, sample size N, and sample $n_1, n_2, \cdots, n_t$, where

$$n_1 + n_2 + \cdots + n_t = N.$$

Let the null hypothesis be that the cell probabilities are $p_1, p_2, \cdots, p_t$, the most interesting case being when $p_1 = p_2 = \cdots = p_t = 1/t$.

The null hypothesis may be tested by any of the following tests among others. Which tests are appropriate will depend largely on what non-null hypotheses are judged to have appreciable initial probabilities.

(A) For the non-null hypothesis we might assume that the cell probabilities are $q_1, q_2, \cdots, q_t$ where $q_1, q_2, \cdots, q_t$ are unknown but are assumed to have what I call a "type II" initial probability density in the simplex (generalised tetrahedron) $q_1 + q_2 + \cdots + q_t = 1$. (By a type II distribution I mean one obtained when sampling from a superpopulation in order to determine an ordinary population.) We should then, by Bayes's theorem, arrive at a factor in favour of the non-null hypothesis provided by the observations $n_1, n_2, \cdots, n_t$. (For the terminology and for the theorem of the weighted average of factors see Good [3], Chapter 6.) This factor would be the weighted average of

$$\left(\frac{q_1}{p_1}\right)^{n_1} \cdots \left(\frac{q_1}{p_t}\right)^{n_t},$$

with weights proportional to the initial probability density. For example, if the initial probability density is taken as proportional to $(p_1 p_2 \cdots p_t)^{\alpha} (\alpha > -1)$ (cf. Perks [20], Good [4], or, for the uniform distribution $\alpha = 0$, Lidstone [18] and Jeffreys [13]), the factor in favour of the non-null hypothesis turns out to be

$$F(\alpha) = \frac{(t\alpha + t - 1)! \prod_{r=1}^{t} (n_r + \alpha)!}{(\alpha!)^t (N + t\alpha + t - 1)! \prod_r p_r^{n_r}}.$$

In accordance with what I called the "Bayes/non-Bayes synthesis" in lectures in Princeton and Chicago in 1955, we could,[1] for some guessed value of α, regard $F(\alpha)$ simply as a statistic and work with its distribution, given only the null hypothesis. [The Bayes/non-Bayes synthesis is the following technique for synthesizing subjective and objective methods in statistics. (i) We use the neo/Bayes-Laplace philosophy in order to arrive at a factor, F, in favour of the non-null hypothesis. For the particular case of discrimination between two simple statistical hypotheses, the factor in favour of a hypothesis is equal to the likelihood ratio, but not in general. The neo/Bayes-Laplace philosophy usually works with inequalities between probabilities, but for definiteness we here assume that the initial distributions are taken as precise, though not necessarily uniform. (ii) We then use F as a statistic and try to obtain its distribution on the null hypothesis, and work with its tail-area probability, P. (iii) Finally we look to see if F lies in the range

$$\left(\frac{1}{30P}, \frac{3}{10P}\right).$$

If it does not lie in this range we think again.]

(B) *The likelihood-ratio test.* (See, for example, Wilks [26].) Minus twice the logarithm of the likelihood ratio is

$$\mu = 2\sum n_r \log_e n_r - 2\sum n_r \log_e p_r - 2N \log_e N,$$

which has, asymptotically, a gamma-variate distribution with $t - 1$ degrees of freedom. For the equiprobable case

$$\mu = 2\sum n_r \log_e n_r + 2N \log_e t - 2N \log_e N.$$

(C) *The chi-squared test* (for numerous references see, for example, the Index of M. G. Kendall [16], Vol. II),

$$\chi^2 = \sum (n_r - Np_r)^2/(Np_r).$$

χ^2 arises as an approximation either to μ or to a constant plus 2 log $F(\alpha)$, whatever the value of α. In fact, for any initial distribution with positive density at $(p_1, p_2, \cdots, p_t)$, the log-factor ("weight of evidence") in favour of the non-null hypothesis, when the null hypothesis is true, is asymptotically of the form

$$\tfrac{1}{2}\chi^2 - K,$$

where K depends only on N, t, $p_1, \cdots, p_t$ and not further on the sampling frequencies $n_1, n_2, \cdots, n_t$. This, to a neo/Bayes-Laplacian, is the real justification for the use of χ^2. A similar argument applies to the use of χ^2 for testing absence of association in contingency tables.

Among the advantages and disadvantages of μ as compared with χ^2 are (i) μ more closely puts the possible samples (for given N, t, $p_1, p_2, \cdots, p_t$) in order of their likelihoods on the null hypothesis, (ii) when tables of $2n \log_e n$ are available, the calculation of μ can be done by additions, subtractions, and table-lookups only, but the calculation is less "well-conditioned" than for χ^2, in the sense that more significant figures must be held, (iii) χ^2 is a simpler mathematical function of the observations and it should be easier to approximate closely to its distribution, given the null hypothesis. (See Sec. 8, where a method is discussed for improving on the usual gamma-variate approximation.)

(D) *Number of zero entries.* Sometimes the (possibly vague) non-null hypothesis has a lot of type II probability density close to regions where several q_r's vanish. In this case a reasonable statistics is the number of zero n_r's. The probability that the number of zero n_r's will be exactly s is

$$\sum\left\{(1-q)^N - \sum_{\mu}(1-q+p_{r_\mu})^N + \sum_{\mu,\nu}^{\mu\neq\nu}(1-q+p_{r_\mu}+p_{r_\nu})^N - \cdots\right\},$$

where

$$q = p_{r_1} + p_{r_2} + \cdots + p_{r_s},$$

and where the outer summation is over all unequal values of $r_1, r_2, \cdots, r_s$. For the equiprobable null hypothesis the above probability reduces to

$$\binom{t}{s}\left\{\left(\frac{t-s}{t}\right)^N - \binom{s}{1}\left(\frac{t-s-1}{t}\right)^N + \cdots + (-1)^{s-1}\binom{s}{s-1}\left(\frac{1}{t}\right)^N\right\}$$
$$= \frac{1}{t^N}\binom{t}{s}\Delta^s 0^N,$$

and is discussed in some detail by Rao and Chakravarti [21].

(E) *Maximum entry,*

$$\max n_r.$$

This statistic would have some application to some work of Guttman [8], as pointed out by Greenwood and Glasgow [7]. The latter authors considered the distribution of both maximum and minimum entries in a binomial distribution and discussed also a special trinomial example. Their methods and results hardly overlap with ours. The distribution of the statistic $\max_r n_r$ can be obtained by means of the saddle-point method and is discussed further in Secs. 4, 7, and 9.

4. Some generating functions, mainly related to the multinomial distribution.

4.1. *Maximum and minimum entries.* Let $P(\text{all } n_r \leqq m)$ be denoted by

$$P(m \mid N, t),$$

which of course depends on $m, N, t, p_1, p_2, \cdots, p_t$. For the equiprobable hypothesis $p_1 = p_2 = \cdots = p_t$ we denote the probability by $P_0(m \mid N, t)$, a function of m, N, and t only. Then it can be at once verified that (for all x)

$$\sum_{N=0}^{\infty}\frac{x^N}{N!}P(m \mid N, t) = \prod_{r=1}^{t}\left(1 + p_r x + \cdots + \frac{p_r^m x^m}{m!}\right), \tag{4.1}$$

and in particular that

(4.2) $$\sum_{N=0}^{\infty} \frac{t^N x^N}{N!} P_0(m \mid N, t) = \left(1 + x + \frac{x^2}{2!} + \cdots + \frac{x^m}{m!}\right)^t.$$

Similarly the probability that the minimum entry is at least m is the coefficient of x^N in

(4.3) $$N! \prod_{r=1}^{t} \left(\frac{p_r x^m}{m!} + \frac{p_r^{m+1} x^{m+1}}{(m+1)!} + \cdots\right),$$

or, for the equiprobable multinomial distribution, in

(4.4) $$\frac{N!}{t^N}\left(\frac{x^m}{m!} + \frac{x^{m+1}}{(m+1)!} + \cdots\right)^t.$$

The above four generating functions are simple generalisations of the one in Proposition XXIII of Whitworth [25].

4.2. *Probability that all n_r are even or all are odd.* It may be noticed, for its entertainment, that the probability that all the n_r's are even, for an equiprobable multinomial distribution, is equal to the coefficient of x^N in

$$\frac{N!}{t^N}\left(1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right)^t,$$

i.e. in

$$\frac{N!}{2^t t^N}\left\{e^{tx} + te^{(t-2)x} + \binom{t}{2} e^{(t-4)x} + \cdots + e^{-tx}\right\},$$

so the probability is

$$2^{-t}\left\{1 + \binom{t}{1}\left(1 - \frac{2}{t}\right)^N + \binom{t}{2}\left(1 - \frac{4}{t}\right)^N + \cdots + \left(1 - \frac{2t}{t}\right)^N\right\}$$
$$\approx 2^{-t+1}(1 + e^{-2N/t})^t$$

if t is large and N is even. Similarly the probability that all n_r's are odd is

$$2^{-t}\left\{1 - \binom{t}{1}\left(1 - \frac{2}{t}\right)^N + \binom{t}{2}\left(1 - \frac{4}{t}\right)^N - \cdots + (-1)^t\left(1 - \frac{2t}{t}\right)^N\right\}$$
$$= \frac{2^{N-t}}{t^N} \nabla^t \left(\frac{1}{2} t\right)^N,$$

and is approximately equal to $2^{-t+1}(1 - e^{-2N/t})^t$ if t is large and $N \equiv t \pmod 2$.

4.3 *Chi-squared for the equiprobable multinomial distribution.* We write, as is customary,

$$\chi^2 = \sum_{r=1}^{t} (n_r - N/t)^2/(N/t)$$

and we have

(4.5) $$\chi^2 = \frac{tS}{N} - N,$$

where $S = \sum_{r=1}^{t} n_r^2$, so that the problem of the distribution of χ^2 is essentially the same as that of S. Now we can at once verify that, M being an integer,

$$\sum_{M,N=0}^{\infty} \frac{x^M y^N t^N}{N!} \Pr(S = M) = \left(\sum_{n=0}^{\infty} \frac{x^{n^2} y^n}{n!} \right)^t. \tag{4.6}$$

In the hope that they may suggest to the reader some improvements in the analysis of this paper, we mention a few facts about the function

$$F(x, y) = \sum_{n=0}^{\infty} \frac{x^{n^2} y^n}{n!}.$$

We have

$$\frac{\partial F(e^\xi, e^\eta)}{\partial \xi} = \frac{\partial^2 F(e^\xi, e^\eta)}{\partial \eta^2},$$

$$| F(e^{i\theta}, y) |^2 = \sum_{n=0}^{\infty} (2y)^n \cos^n (n\theta)/n!,$$

$$F(e^{-\alpha}, y) = \frac{1}{\sqrt{2\pi\alpha}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\alpha} + y e^{iu\sqrt{2}}} \, du, \qquad (R\alpha > 0).$$

Finally we mention that the function $F(e^{\pi i \sqrt{2}}, y)$ is discussed by Nassif [19] and Tims [23].

4.4. *Chi-squared for a contingency table.* For an r by s contingency table $\{n_{ij}\}(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s; \sum_{i,j} n_{ij} = N, \sum_j n_{ij} = n_{i.}, \sum_i n_{ij} = n_{.j})$, the probability of the table, given the borders, and assuming no association, is, as is well known,

$$\frac{\prod_{ij} n_{i.}! n_{.j}!}{N! \prod_{i,j} n_{ij}!}.$$

Now

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - n_{i.} n_{.j}/N)^2}{n_{i.} n_{.j}/N} = N(S - 1),$$

where

$$S = \sum_{i,j} \frac{n_{ij}^2}{n_{i.} n_{.j}}.$$

The problem of the distribution of χ^2 is equivalent to that of S, and

$$\Pr(S = M) = \text{coefficent of } z^M \prod_{i,j} x_i^{n_{i.}} y_j^{n_{.j}} \text{ in}$$

$$\frac{\prod_{i,j} n_{i.}! n_{.j}!}{N!} \prod_{i,j} \sum_{n=0}^{\infty} \frac{x_i^n y_j^n z^{n^2/(n_{i.} n_{.j})}}{n!}. \tag{4.7}$$

The moment generating function of S is therefore the coefficient of $\prod_{i,j} x_i^{n_{i\cdot}} y_j^{n_{\cdot j}}$ in

$$\frac{\prod_{i,j} n_{i\cdot}!\, n_{\cdot j}!}{N!} \prod_{i,j} \sum_{n=0}^{\infty} \frac{x_i^n y_j^n e^{n^2 u/(n_{i\cdot} n_{\cdot j})}}{n!}.$$

5. Methods of evaluating coefficients in generating functions.

5.1. *Continuous methods.* A coefficient in a power series may be expressed as a contour integral by means of Cauchy's formula, and then this integral may be expanded into an asymptotic series by a saddle-point method. In the present paper examples are given in Secs. 7 and 8. For continuous random variables with probability densities the analogous process is the use of the Fourier or Laplace transform instead of Cauchy's integral.

5.2. *Discrete methods.* Let

$$g(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}.$$

Then if u is a positive integer we have

$$\frac{1}{u} \sum_{r=0}^{u-1} g\left(\frac{2\pi r}{u}\right) = \sum_{n=-\infty}^{\infty} c_{nu}. \tag{5.1}$$

In particular, if $c_n = 0$ whenever $| n | \geqq m$, then

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta)\, d\theta = c_0 = \frac{1}{u} \sum_{r=0}^{u-1} g\left(\frac{2\pi r}{u}\right) \tag{5.2}$$

whenever $u \geqq m$. (Compare D. G. Kendall [14], and Good [5].) Given a generating function that happens to be a polynomial, $h(x)$, we can extract the coefficient of x^N by using (5.2) with $g(\theta) = e^{-Ni\theta} h(e^{i\theta})$. An example will be given in Sec. 9, and it should be noticed that the method gives an exact formula (from which an asymptotic formula can sometimes be deduced). For continuous random variables the analogous process would be the use of Poisson's summation formula: see, for example, Krishnan [17].

When the generating function is an infinite power series we can make similar use of (5.1) provided that the series on the right is utterly dominated by its largest term. A potential example is given in Sec. 10.

6. The general asymptotic formulae.

6.0. In this section we give theorems in a convenient form for applications to probability and statistics, concerned with the asymptotic expension of multiple integrals containing a large parameter. Three terms are given for single integrals, two for double integrals, and one for multiple integrals. Only the first two theorems are applied in the present paper, but the multidimensional theorem seems worth stating since it puts the two-dimensional one into proper perspective.

6.1. *The asymptotic expansion of a single integral.*

THEOREM 6.1. *Let*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$$

be a power-series or polynomial with non-negative real coefficients and a non-vanishing open domain of convergence (an annulus, or the inside or outside of a circle). Suppose that the suffixes r at which $a_r > 0$ do not all have a common factor greater than 1. (This condition is given, though incidentally, by Daniels [2], p. 646. It is clearly necessary and its sufficiency will follow from the remarks following Theorem 6.3. The condition can always be forced by a change in the variable of the generating function.) Let the coefficient of z^N in $(f(z))^t$ be $c(N, t)$. If $c(N, t) \neq 0$, then there is a unique non-negative real solution, ρ, of the equation

$$t\rho \frac{d}{d\rho} f(\rho) = Nf(\rho); \tag{6.1}$$

and, if in addition N/t is held inside a constant interval, then

$$c(N, t) \sim \frac{[f(\rho)]^t}{\sigma\rho^N\sqrt{2\pi t}} \cdot \left\{1 + \frac{1}{24t}(3\lambda_4 - 5\lambda_3^2) + (1/1152t^2) \cdot (168\lambda_3\lambda_5 + 385\lambda_3^4 - 630\lambda_3^2\lambda_4 - 24\lambda_6 + 105\lambda_4^2) + \cdots\right\} \tag{6.2}$$

uniformly as $t \to \infty$, where

$$\lambda_s = \lambda_s(\rho) = \kappa_s(\rho)/\sigma^s \tag{6.3}$$

where

$$\sigma = \sqrt{\kappa_2(\rho)}, \tag{6.4}$$

and

$$\kappa_s(\rho) = \left(\frac{\partial}{\partial u}\right)^s (\log f(\rho e^u))\,|_{u=0}, \qquad (s = 0, 1, 2, \cdots). \tag{6.5}$$

If we write $\rho = e^\xi$, we may replace (6.5) by

$$\kappa_s(\rho) = \left(\frac{\partial}{\partial \xi}\right)^s \log f(e^\xi)\,|_{\xi=\log\rho}. \tag{6.6}$$

Similarly if $f(z) = \int_{-\infty}^{\infty} a(r)z^r\,dr$, where

(a) $a(r) \geqq 0$ for all real r, and is continuous,

(b) the integral is convergent for some non-vanishing open interval of positive values of z, then

$$(f(z))^t = \int_{-\infty}^{\infty} c(N, t)z^N\,dN,$$

where $c(N, t)$ is still given by (6.2) for positive [negative] N, and where the remaining conclusions are formally the same as in the discrete case. The conclusions of the theorem (both in the discrete and continuous cases) may be simultaneously true for positive and negative N.

The above theorem can be proved as in Daniels [2] and differs from Daniels's form only in (i) that we do not insist on the condition $f(1) = 1$, and (ii) that we have calculated the third term of the asymptotic series. More terms could be worked out on an electronic computer programmed to do algebra.

Since the theorem is so similar to the form given by Daniels we shall here content ourselves with some of the formal details leading to the extra term. We may suppose without real loss of generality that $f(\rho) = 1$. Write $\kappa_r(\rho) = \kappa_r$ for short. Then

$$\begin{aligned} c(N,t) &= \frac{1}{2\pi i}\oint (f(z))^t z^{-N-1}\,dz \\ &= \frac{1}{\rho^N 2\pi}\int_{-\pi}^{\pi} (f(\rho e^{i\theta}))^t e^{-Ni\theta}\,d\theta \\ &= \frac{1}{\rho^N 2\pi}\int_{-\pi}^{\pi} \exp t\left\{-\tfrac{1}{2}\kappa_2\theta^2 - \frac{i\kappa_3\theta^3}{6} + \cdots\right\} d\theta \\ &= \frac{1}{\rho^N 2\pi\sqrt{t}}\int_{-\pi\sqrt{t}}^{\pi\sqrt{t}} e^{-\frac{1}{2}\kappa_2\varphi^2} \exp\left(\frac{\kappa_4\varphi^4}{24t} - \frac{\kappa_6\varphi^6}{720t^2} + \cdots\right) \\ &\qquad\qquad \cdot \cos\left(\frac{\kappa_3\varphi^3}{6t^{1/2}} - \frac{\kappa_5\varphi^5}{120t^{3/2}} + \cdots d\varphi\right) \end{aligned}$$

and (6.2) follows from the formula

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} x^{2n}\,dx = 1.3.5.\cdots(2n+1).$$

(Some of the above algebra can be done with the help of Kendall [16], I, formula (3.30), with his $\kappa_2 = 0$, since this formula gives the expansion of the exponential of a power series.) As a check of the above theorem consider the case $a_r = e^{-N/t}(N/t)^r/r!$. Then we find that the theorem correctly gives

$$\frac{1}{N!} \sim \frac{e^N}{N^{N+\frac{1}{2}}\sqrt{2\pi}}\left(1 - \frac{1}{12N} + \frac{1}{288N^2} + \cdots\right).$$

A further check of the theorem can be obtained by applying it to a classical problem in the theory of numbers, namely the enumeration of ways of expressing a positive integer N as the sum of t squares of numbers $0, \pm 1, \pm 2, \cdots$ (different orders counting as different representations). For a detailed discussion of this problem see Hardy [9], Chapter IX. Here $c(N, t) = r_t(N)$ in Hardy's notation, and

$$\sum_{N=0}^{\infty} r_t(N)x^N = (\vartheta(x))^t,$$

where

$$\vartheta(x) = \sum_{n=-\infty}^{\infty} x^{n^2} \qquad (|x| < 1).$$

The equation for ρ becomes

$$\sum_{-\infty}^{\infty} n^2 \rho^{n^2} = \frac{N}{t} \sum_{-\infty}^{\infty} \rho^{n^2}. \tag{6.7}$$

Now

$$f(\rho) = \sum_{-\infty}^{\infty} \rho^{n^2} = \sqrt{\frac{\pi}{\log 1/\rho}} \sum_{-\infty}^{\infty} e^{-n^2\pi^2/\log\frac{1}{\rho}}$$

$$\approx \sqrt{\frac{\pi}{\log 1/\rho}}, \qquad \text{if } \rho \text{ is near } 1.$$

If ρ is near to 1 we may approximate the derivatives of $\log f(\rho)$ by those of

$$-\tfrac{1}{2} \log \log \frac{1}{\rho}.$$

We find that

$$\log \frac{1}{\rho} \approx \frac{t}{2N}$$

if N/t is large, that

$$\sigma \approx \frac{N}{t} \sqrt{2},$$

and that

$$\kappa_r(\rho) \approx \frac{(r-1)!}{2} \left(\frac{2N}{t}\right)^r \qquad (r = 1, 2, \cdots);$$

$$\lambda_3 \approx 2\sqrt{2}, \qquad \lambda_4 \approx 12, \qquad \lambda_5 \approx 48\sqrt{2}, \qquad \lambda_6 \approx 480;$$

and Theorem 6.1 gives

$$r_t(N) \approx \frac{1}{2N} \sqrt{\frac{t}{\pi}} \left(\frac{2\pi N e}{t}\right)^{\frac{1}{2}t} \left(1 - \frac{1}{6t} + \frac{1}{72t^2} + \cdots\right).$$

Clearly what we have here is no better than the elementary result

$$r_t(N) \sim \frac{\pi^{\frac{1}{2}t} N^{\frac{1}{2}t-1}}{\Gamma(\frac{1}{2}t)} \quad \text{as} \quad N \to \infty,$$

though once again the example acts as a check of Theorem 6.1. It would be interesting to see what the theorem would give if N/t were not assumed to be large. The equation (6.7) could be solved iteratively and the result of the theorem could be compared with known results; for example, the case $t = 24$ is treated in detail by Hardy [9].

On the whole is seems unlikely that the present method could compete in elegance with the classical theory in which, instead of the saddle-point method, a so-called "singular series" is obtained by using a contour that comes very close to the unit circle. It may be that there is scope for the method of singular series in statistical problems.

6.2. *The asymptotic expansion of a double integral.* (Compare Hsu [11]; Copson [1], Sec. 5. These authors do not work out the second term of the expansion.)

THEOREM 6.2. *Let*

$$f(x, y) = \sum_{r=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a(r, s) x^r y^s$$

(or $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(r, s) x^r y^s \, dx \, dy$, *where* $a(r, s)$ *is continuous), where the conditions* (6.14), (6.15), *and* (6.16) *of the (multidimensional) Theorem* 6.3 *are satisfied with* $l = 2$. *Let the coefficient of* $x^M y^N$ *in* $(f(x, y))^t$ *be* $c(M, N, t)$, *i.e.,*

$$(f(x, y))^t = \sum_{M=-\infty}^{\infty} \sum_{N=-\infty}^{\infty} c(M, N, t) \, x^M y^N$$

$$\left(\text{or} \quad (f(x, y))^t = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(M, N, t) \, x^M \, y^N \, dM \, dN\right),$$

where the conditions (6.17) *to* (6.19) *of Theorem* 6.3 *are satisfied with* $l = 2$. *Then there is at most one pair of non-negative real numbers* ρ, ρ' *such that*

$$t\rho \frac{\partial}{\partial \rho} f(\rho, \rho') = M f(\rho, \rho'), \tag{6.8}$$

$$t\rho' \frac{\partial}{\partial \rho'} f(\rho, \rho') = N f(\rho, \rho'), \tag{6.9}$$

(If (ρ, ρ') *is a boundary point of the domain of convergence, then* f *and its derivatives may be interpreted as limits from within the domain. A similar understanding applies throughout this paper.)*
and if there is such a pair then

$$\begin{aligned} c(M, N, t) \sim \frac{[f(\rho, \rho')]^t}{2\pi t \rho^M \rho'^N \sqrt{\Delta}} \Big\{ 1 + \frac{1}{24t} [& 3\lambda_{40} - 12\lambda_{31} \alpha + 6\lambda_{22}(1 + 2\alpha^2) \\ & - 12\lambda_{13} \alpha + 3\lambda_{04} - 5\lambda_{30}^2 - 9\lambda_{21}^2(1 + 4\alpha^2) - 9\lambda_{12}^2(1 + 4\alpha^2) \\ & - 5\lambda_{03}^2 + 30\lambda_{30} \lambda_{21} \alpha - 6\lambda_{30} \lambda_{12}(1 + 4\alpha^2) + 2\lambda_{30} \lambda_{03} \alpha(3 + 2\alpha^2) \\ & + 18\lambda_{21} \lambda_{12} \alpha(3 + 2\alpha^2) - 6\lambda_{21} \lambda_{03}(1 + 4\alpha^2) \\ & + 30\lambda_{12} \lambda_{02} \alpha] + \cdots \Big\}, \end{aligned} \tag{6.10}$$

where

$$\lambda_{r,s} = \frac{\kappa_{rs} \, \kappa_{02}^{\frac{1}{2}r} \, \kappa_{20}^{\frac{1}{2}s}}{\Delta^{\frac{1}{2}(r+s)}}, \qquad \alpha = \frac{\lambda_{11}}{\sqrt{(\lambda_{20} \lambda_{02})}}, \tag{6.11}$$

where

$$\Delta = \kappa_{20}\kappa_{02} - \kappa_{11}^2, \tag{6.12}$$

where

$$\kappa_{rs} = \kappa_{r,s}(\rho, \rho') = \left(\frac{\partial}{\partial\xi}\right)^r \left(\frac{\partial}{\partial\eta}\right)^s \log f(e^\xi, e^\eta)\Big|_{\substack{\xi=\log\rho \\ \eta=\log\rho'}}. \tag{6.13}$$

Here again we merely give some of the formal details of the proof. These details should suffice, when combined with the references cited, together with the remarks following Theorem 6.3.

There is no real loss of generality in assuming that $f(\rho, \rho') = 1$. Then

$$c(M, N, t) = \frac{1}{\rho^M \rho'^N 4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp t\left\{ -\frac{1}{2}\kappa_{20}\theta^2 - \kappa_{11}\theta\varphi - \frac{1}{2}\kappa_{02}\varphi^2 - \frac{i}{6}\kappa_{30}\theta^3 - \cdots \right\} d\theta\, d\varphi$$

$$= \frac{1}{\rho^M \rho'^N 4\pi^2} \int_{-\pi\sqrt{t}}^{\pi\sqrt{t}} \int_{-\pi\sqrt{t}}^{\pi\sqrt{t}} e^{-\frac{1}{2}(\kappa_{20}\theta^2 + 2\kappa_{11}\theta\varphi + \kappa_{02}\varphi^2)} \exp\left\{1 + \frac{1}{24t}(\kappa_{40}\theta^4 + 4\kappa_{31}\theta^3\varphi + 6\kappa_{22}\theta^2\varphi^2 + 4\kappa_{13}\theta\varphi^3 + \kappa_{04}\varphi^4) + \cdots\right\}$$
$$\cos\left\{\frac{1}{6t^{\frac{1}{2}}}(\kappa_{30}\theta^3 + 3\kappa_{21}\theta^2\varphi + 3\kappa_{12}\theta\varphi^2 + \kappa_{03}\varphi^3) + \cdots\right\} d\theta\, d\varphi.$$

If we now define $\bar\sigma_1^2$, $\bar\sigma_2^2$ and $\bar\rho$ by the equations

$$\bar\sigma_1^2 = \frac{\kappa_{02}}{\Delta}, \qquad \bar\sigma_2^2 = \frac{\kappa_{20}}{\Delta}, \qquad \bar\rho = \frac{-\kappa_{11}}{(\kappa_{20}\,\kappa_{02})^{\frac{1}{2}}},$$

where $\Delta = \kappa_{02}\kappa_{20} - \kappa_{11}^2$, we may apply the formulae for the moments of a bivariate normal distribution as given, for example, by Kendall [16], Sec. 3.29 and Exercise 3.15, and we formally obtain the result (6.10). In order to check the algebra it may be observed that, when $\alpha = 1$, the sum of the moduli of the first five coefficients of the λ's is 48 and the sum of the rest is 320, while the algebraic sums are both zero. These facts can be inferred from the above argument without going through all the details.

An application of Theorem 6.2 is given in Sec. 8. In this application it is found that the coefficient of $1/(24t)$ is very ill-conditioned. It may be possible to write it in a well-conditioned form.

6.3. *The asymptotic expansion of a multiple integral.* (Compare Hsu [12], Rooney [22].)

THEOREM 6.3. *Let*

$$f(x_1, x_2, \cdots, x_l) = \sum_{r_1, r_2, \cdots, r_l} a(r_1, \cdots, r_l) x_1^{r_1} \cdots x_l^{r_l}$$
$$\left(\text{or} \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} a(r_1, \cdots, r_l) x_1^{r_1} \cdots x_l^{r_l}\, dr_1 \cdots dr_l\right),$$

where $a(r_1, \cdots, r_l)$ is continuous), where the summation is over all integral lattice-points in l-dimensional space, and where

$$a(r_1, r_2, \cdots, r_l) \geqq 0; \tag{6.14}$$

(6.15) *the series (integral) is convergent in some non-vanishing open l-dimensional domain;*

(6.16) *there exist positive integers $R_{10}, R_{20}, \cdots, R_{l0}$ such that every point $R_1, R_2, \cdots, R_l$ with $R_1\epsilon_1 > R_{10}, R_2\epsilon_2 > R_{20}, \cdots, R_l\epsilon_l > R_{l0}$, where each ϵ is either $+1$ or -1, can be expressed as a linear combination of points $(r_1, r_2, \cdots, r_l)$ for which $a(r_1, r_2, \cdots, r_l) > 0$, the coefficients in these linear combinations being positive integers. In other words the suffixes corresponding to positive coefficients in f span all points sufficiently far from the origin in at least one "octant" (or 2^l-ant). (For better understanding the reader may take $\epsilon_j = 1$ for all j.) There may be as many as 2^l octants for which this condition is valid. (In the "continuous" case of the theorem this condition does not require explicit mention.)*

Let the coefficient of $x_1^{M_1}x_2^{M_2} \cdots x_l^{M_l}$ in $(f(x_1, x_2, \cdots, x_l))^t$ be

$$c(M_1, M_2, \cdots, M_l; t),$$

i.e.,

$$(f(x_1, x_2, \cdots, x_l))^t = \sum_{M_1=-\infty}^{\infty} \cdots \sum_{M_l=-\infty}^{\infty} c(M_1, \cdots, M_l)x_1^{M_1} \cdots x_l^{M_l}$$

$$\left(\text{or} \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} c(M_1, \cdots, M_l)x_1^{M_1} \cdots x_l^{M_l}\, dM_1 \cdots dM_l\right).$$

Suppose further that we restrict our attention to a class of values of

$$(M_1, M_2, \cdots, M_l)$$

for which

(6.17) $M_j/t (j = 1, 2, \cdots, l)$ all lie in fixed, not necessarily finite, intervals;

$$\operatorname{sgn} M_j = \epsilon_j \qquad (j = 1, 2, \cdots, l); \tag{6.18}$$

(6.19) $c(M_1, M_2, \cdots, M_l; t) \neq 0$ at any point satisfying the above conditions, if t is large enough.

Then there is at most one ordered set of l non-negative real numbers $\rho_1, \rho_2, \cdots, \rho_l$ such that

$$t\rho_j \frac{\partial}{\partial \rho_j} f(\rho_1, \cdots, \rho_l) = M_j f(\rho_1, \cdots, \rho_l) \qquad (j = 1, 2, \cdots, l), \tag{6.20}$$

and if there is such a set of l numbers, then uniformly

$$c(M_1, M_2, \cdots M_l; t) \sim \frac{[f(\rho_1, \rho_2, \cdots, \rho_l)]^t}{(2\pi t)^{\frac{1}{2}l} \rho_1^{M_1} \cdots \rho_l^{M_l} \sqrt{\Delta}}, \tag{6.21}$$

where

$$\Delta = \det \{\kappa_{11}^{jk}(\rho_1, \cdots, \rho_l)\}, \tag{6.22}$$

where

$$(6.23)\qquad \kappa_{11}^{jk}(\rho_1, \cdots, \rho_l) = \rho_j \frac{\partial}{\partial \rho_j}\left(\rho_k \frac{\partial}{\partial \rho_k} \log f(\rho_1, \cdots, \rho_l)\right)$$

even if $j = k$. Or we can write

$$(6.24)\qquad \kappa_{11}^{jk}(\rho_1, \cdots, \rho_l) = \frac{\partial^2}{\partial \xi_j \, \partial \xi_k} \log f(e^{\xi_1}, \cdots, e^{\xi_k}) \Big|_{\substack{\xi_j = \log \rho_j \\ (j=1,\cdots,l)}} .$$

In other words Δ is the Hessian of $\log f(e^{\xi_1}, \cdots, e^{\xi_l})$ at

$$(\xi_1, \cdots, \xi_l) = (\log \rho_1, \cdots, \log \rho_l).$$

Notice that

$$f^* = f(e^{\xi_1}, \cdots, e^{\xi_l}) e^{-(M_1 \xi_1 + \cdots + M_l \xi_l)/t},$$

which is essentially of the form of a Laplace transform of a non-negative function, is an analytic convex function of $(\xi_1, \xi_2, \cdots, \xi_l)$ in its real domain of convergence, and so also is $\log f^*$. (Cf. Doetsch [2A], p. 58.) It can be seen to be *strictly* convex if the points ("basis vectors") at which $a(r_1, r_2, \cdots, r_l) > 0$ span a genuinely l-dimensional space, which they do in virtue of condition (6.16). (It is also strictly convex in any linear manifold that belongs to the boundary of the domain of convergence.) Hence the Hessian of $\log f^*$ (which is equal to Δ) is strictly positive at points of the real domain of convergence of f, and stationary points of f^*, even if they are on the boundary of the domain, are necessarily minimum points. There cannot be more than one stationary point. Certainly f^* attains its minimum but this may be on the boundary of the domain. It follows that the solution of (6.20) is unique if it exists, and it will exist if f^* attains its minimum at an interior point. It may also exist if the minimum is on the boundary, provided that the minimum is a stationary point as in an example considered in Sec. 8 in relation to the distribution of χ^2. In any practical problem if the equations (6.20) can be solved there are no further difficulties.

It can be seen that condition (6.16) is equivalent to the statement that every point in the l-dimensional lattice can be expressed linearly, with integral coefficients (not necessarily positive), in terms of the "basis vectors"

$$(r_1, r_2, \cdots, r_l)$$

for which $a(r_1, r_2, \cdots, r_l) > 0$. If this condition were not satisfied, then at least one of the vectors $(1, 0, 0, \cdots, 0)$, $(0, 1, 0, \cdots, 0)$, $(0, 0, 1, \cdots, 0)$, etc., would not be "spanned," say $(1, 0, 0, \cdots, 0)$ for definiteness. In this case there would be a smallest positive integer r_0 such that all multiples of $(r_0, 0, 0, \cdots, 0)$ would be spanned and no other points of the form $(r, 0, 0, \cdots, 0)$. Then there would be at least r_0 values of θ_1, not congruent modulo 2π, at which

$$f(\rho_1 e^{i\theta_1}, \rho_2 e^{i\theta_2}, \cdots, \rho_l e^{i\theta_l})$$

would be of maximum modulus, i.e., at least r_0 equally important saddlepoints, and the asymptotic formula (6.21) would need modification.

The remaining details of rigour in the proof of Theorem 6.3 may be supplied along the lines of Daniels [2], whose use of Lagrange's expansion must be now replaced by the multidimensional form that is given in Sec. 104 of Goursat [6] (and attributed to Laplace).

If in Theorem 6.3 the power f^t were replaced by the product of t distinct power series, and if moreover the second term in the asymptotic expansion were obtained then we should have a theorem that could be used in conjunction with (4.7) for approximation to the distribution of χ^2 for contingency tables. If only the first term of the asymptotic expansion were available we should merely arrive at the familiar gamma-variate approximation.

7. Asymptotic expansion of $P_0(m \mid N, t)$. We pointed out in Sec. 4.1 that $P_0(m \mid N, t)$ is equal to the coefficient of x^N in

$$\frac{N!}{t^N}\left(1 + x + \frac{x^2}{2!} + \cdots + \frac{x^m}{m!}\right)^t.$$

We may therefore make use of Theorem 6.1. with

$$f(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^m}{m!}.$$

Equation (6.1) becomes

$$\rho + \frac{\rho^2}{1!} + \cdots + \frac{\rho^m}{(m-1)!} = \frac{N}{t}\left(1 + \rho + \frac{\rho^2}{2!} + \cdots + \frac{\rho^m}{m!}\right), \tag{7.1}$$

which, when N and t are numerically assigned, can be solved by any method for the numerical solution of algebraic equations. It is then a straightforward calculation to apply Theorem 6.1, and it could be done on a general-purpose computer for any specified values of t, N, and m.

As a detailed example take $m = 2$ and $N = t$. Then

$$\rho = \sqrt{2}$$

and it can be shown that

$$\begin{aligned}\sum_{r=0}^{\infty} \frac{\kappa_r u^r}{r!} &= \log f(\rho e^u) \\ &= \log f(\rho) + u + \log\left[1 + (4 - 2\sqrt{2}) \sinh^2 \frac{u}{2}\right].\end{aligned}$$

Hence

$$\kappa_3 = \kappa_5 = \cdots = 0,$$

$$\kappa_1 = 1, \qquad \kappa_2 = 2 - \sqrt{2}, \qquad \kappa_4 = -16 + 11\sqrt{2}, \qquad \kappa_6 = 512 - 361\sqrt{2},$$

$$\lambda_4 = \tfrac{1}{2}(-4 + \sqrt{2}), \qquad \lambda_6 = \tfrac{1}{2}(33 - 13\sqrt{2}),$$

TABLE 1

t	One term	Two terms	Three terms	$P_0(2 \mid t, t)$
1	1.160	1.070	1.058	$1.\dot{0}$
2	1.031	0.990	0.988	$1.\dot{0}$
3	0.9153	0.8914	0.8904	$0.888\dot{8}$
4	0.8129	0.79703	0.79652	$0.796875\dot{0}$
5	0.7220	0.71326	0.7104105	$0.7104\dot{0}$
10	0.3990	0.39585	0.3958103	$0.395811{:}36\dot{0}$

and

$$P_0(2 \mid t, t) \sim \frac{1}{\sqrt{2-\sqrt{2}}}\left(\frac{1+\sqrt{2}}{e}\right)^t \cdot \left\{1 - \frac{8-3\sqrt{2}}{48t} - \frac{96\sqrt{2}-113}{2304t^2} + \cdots\right\}. \tag{7.2}$$

In Table 1 the results of taking one, two and three terms of (7.2) are given in the second, third, and fourth columns, while the last column gives the exact value of $P_0(2 \mid t, t)$. It seems fair to say that, when using formula (7.2), we may regard 4 as a large number and 10 as very large indeed.

When N is not necessarily equal to t, the first term of the asymptotic formula is

$$P_0(2 \mid N, t) \sim \left(\frac{N}{t}\right)^{N+\frac{1}{2}} \frac{2+\sqrt{2}}{(e\sqrt{2})^N\sqrt{2-\sqrt{2}}} \exp\left\{-\frac{(t-N)^2}{2(2-\sqrt{2})t}\right\}. \tag{7.3}$$

For $t = 8$ we have the following numerical values.

N	2	4	6	8	10	12	14
(i)	1.4	.988	.773	.506	.239	.074	.014
(ii)	1.0	.943	.769	.501	.237	.070	.010

where row (i) is obtained from (7.3) and row (ii) is the value of $P_0(2 \mid N, t)$ correct to three decimal places, computed directly from the generating function.

8. Asymptotic expansion for the distribution of chi-squared. It was pointed out in Sec. 4.3 that for the equiprobable multinomial distribution,

$$\chi^2 = \frac{tS}{N} - N, \quad \text{where} \quad S = \sum_{r=1}^{t} n_r^2, \tag{8.1}$$

and that

$$\Pr(S = M) = \text{coefficient of } x^M y^N \text{ in } N!\, t^{-N} (f)^t, \tag{8.2}$$

where

$$f(x, y) = \sum_{n=0}^{\infty} \frac{x^{n^2} y^n}{n!}.$$

The equations for ρ and ρ', in Theorem 6.2, are

$$\sum_0^\infty \frac{n^2\rho^{n^2}\rho'^n}{n!} = \frac{M}{t}\sum_0^\infty \frac{\rho^{n^2}\rho'^n}{n!}, \tag{8.3}$$

$$\sum_0^\infty \frac{n\rho^{n^2}\rho'^n}{n!} = \frac{N}{t}\sum_0^\infty \frac{\rho^{n^2}\rho'^n}{n!}. \tag{8.4}$$

When M is a possible value for S these two equations have a unique solution, and this solution could be obtained by means of an iterative process on a general-purpose computer. (Each equation can be shown to determine ρ uniquely given ρ' and conversely.)

As an example that can be worked out by hand calculation, we consider the special case

$$N = t, \qquad M = 2t, \qquad \chi^2 = t,$$

when

$$\rho = \rho' = 1$$

($\rho = 1$ is on the boundary of convergence of f, but this does not affect the validity of Theorem 6.2.) We have

$$f(e^\xi, e^\eta) = \sum_{n=0}^\infty \frac{e^{n^2\xi+n\eta}}{n!} = e\sum_{r,s}^{0,1,2,\cdots} \frac{\mu'_{rs}\xi^r\eta^s}{r!s!}, \quad \text{say,}$$

where

$$\mu'_{rs} = e^{-1}\sum_{n=0}^\infty \frac{n^{2r+s}}{n!} = b_{2r+s},$$

where

$$e^{e^x} = 1 + e^x + \frac{e^{2x}}{2!} + \cdots = e\left(b_0 + b_1x + \frac{b_2x^2}{2!} + \cdots\right);$$

$$b_r = b_{r-1} + (r-1)b_{r-2} + \binom{r-1}{2}b_{r-3} + \cdots + b_0,$$

$$b_0 = 1, \quad b_1 = 1, \quad b_2 = 2, \quad b_3 = 5, \quad b_4 = 15,$$

$$b_5 = 52, \quad b_6 = 203, \quad b_7 = 877, \quad b_8 = 4140.$$

Write μ_{rs} for the product moments about the mean of the (artificial) distribution with probability generating function $e^{-1}f(x, y)$, and we find by using known relationships between bivariate moments, moments about means and cumulants (see, for example, Kendall [16] Sec. 3.29 and exercise 3.15, and Kendall [15]),

$$\mu_{rs} = \sum_{j,k}\binom{r}{j}\binom{s}{k}(-\mu'_{10})^{r-j}(-\mu'_{01})^{s-k}\mu'_{jk}$$

$$= e^{-1}2^r\sum_{n=0}^\infty \frac{1}{n!}\left(\frac{n^2}{2} - 1\right)^r (n-1)^s.$$

Hence

$\mu_{10} = 0, \quad \mu_{01} = 0, \quad \mu_{20} = 11, \quad \mu_{11} = 3, \quad \mu_{02} = 1;$

$\mu_{30} = 129, \quad \mu_{21} = 25, \quad \mu_{12} = 5, \quad \mu_{03} = 1;$

$\mu_{40} = 2828, \quad \mu_{31} = 488, \quad \mu_{22} = 90, \quad \mu_{13} = 18, \quad \mu_{04} = 4;$

$\kappa_{00} = 1, \quad \kappa_{20} = 11, \quad \kappa_{11} = 3, \quad \kappa_{02} = 1;$

$\kappa_{30} = \mu_{30} = 129, \quad \kappa_{21} = \mu_{21} = 25, \quad \kappa_{12} = \mu_{12} = 5, \quad \kappa_{03} = \mu_{03} = 1;$

$\kappa_{40} = \mu_{40} - 3\mu_{20}^2 = 2465, \quad \kappa_{31} = \mu_{31} - 3\mu_{20}\mu_{11} = 389,$

$\kappa_{22} = \mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}^2 = 61, \quad \kappa_{13} = \mu_{13} - 3\mu_{02}\mu_{11} = 9,$

$\kappa_{04} = \mu_{04} - 3\mu_{02}^2 = 1.$

$\Delta = \kappa_{02}\kappa_{20} - \kappa_{11}^2 = 2, \quad \alpha = 3/\sqrt{11}.$

$\lambda_{20} = 11/2, \quad \lambda_{11} = (3\sqrt{11})/2, \quad \lambda_{02} = 11/2;$

$\lambda_{30} = 129/(2\sqrt{2}), \quad \lambda_{21} = (25\sqrt{11})/(2\sqrt{2}), \quad \lambda_{12} = 55/(2\sqrt{2}),$

$\lambda_{03} = (11\sqrt{11})/(2\sqrt{2});$

$\lambda_{40} = 2465/4, \quad \lambda_{31} = (389\sqrt{11})/4,$

$\lambda_{22} = 671/4, \quad \lambda_{13} = (99\sqrt{11})/4, \quad \lambda_{04} = 121/4.$

Theorem 6.2 now shows that

$$\text{(8.5)} \qquad \Pr\,(\chi^2 = t \mid N = t) = \frac{1}{2\sqrt{\pi t}}\left(1 + \frac{1}{6t} + \cdots\right).$$

The coefficient of $1/(24t)$ in Theorem 6.2 reduces to 4 in this example, although one of its terms is over 30,000. The mere fact that (8.5) looks sensible is therefore quite a good check of the arithmetic and algebra. It is important to remember for future applications that the coefficient of $1/t$ in Theorem 6.2 is liable to be ill-conditioned, especially for machine programming.

It seems likely that the application of the theorem to χ^2 would give better results for its cumulative distribution than for the individual probabilities $\Pr\,[\chi^2 = (tM/N) - N]$. This opinion is supported by the earlier discussion of the lattice-point problem. When t is small, say $t = 2$, the first term, c_N say, of the asymptotic formula for the lattice-point problem is misleading since no prime of the form $4n + 3$ can be expressed as the sum of two squares; but $c_1 + c_2 + \cdots + c_N$ give a good approximation to the number of lattice points in the circle $x^2 + y^2 \leq N$.

9. Some exact formulae for $P_0(2 \mid N, t)$. In order to illustrate the "discrete method" of section 5.2 we now consider the probability $P_0(2 \mid N, t)$ in more detail.

We have (with $\rho = \sqrt{2}$), from (4.2),

$$P_0(2 \mid N, t) = \frac{N!}{2^t t^N} \cdot \frac{1}{2\pi\rho^N} \int_{-\pi}^{\pi} \frac{(2e^{-i\theta} + 2\rho + 2e^{i\theta})^t}{e^{(N-t)i\theta}}\, d\theta.$$

We can apply (5.2) with

$$u > t + |N - t|,$$

and we obtain the following exact finite series.

$$P_0(2 \mid N, t) = \frac{N!(2+\sqrt{2})^t}{ut^N 2^{\frac{1}{2}N}}$$

$$(9.1) \qquad \left\{1 + 2\sum_{r=1}^{\left[\frac{u-1}{2}\right]} \cos\frac{2\pi r(t-N)}{u}\left\{1 - (4 - 2\sqrt{2})\sin^2\frac{\pi r}{u}\right\}^t + (-1)^N(3 - 2\sqrt{2})^t\epsilon\right\},$$

where

$$\epsilon = \begin{cases} 0 & \text{if} \quad u \text{ is odd} \\ 1 & \text{if} \quad u \text{ is even.} \end{cases}$$

It follows that

$$(9.2) \qquad \frac{P_0(2 \mid 2t - N, t)}{P_0(2 \mid N, t)} = \frac{(2t-N)!}{t^{2(t-N)}N!2^{t-N}},$$

a formula that also follows directly from the fact that $t^N P_0(2 \mid N, t)2^{\frac{1}{2}N}/N!$ is the coefficient of z^N in $(1 + \sqrt{2}z + z^2)^t$, which is equal to that of z^{2t-N} by symmetry.

Some further combinatorial formulae, which we give here without proof, are

$$(9.3) \qquad P(1 \mid N, t) = \frac{(t-1)!}{(t-N)!t^{N-1}}$$

$$(9.4) \qquad P_0(2 \mid N, t) = \frac{N!t!}{t^N}\sum_{s=0}^{\infty}\frac{1}{(N-2s)!(t-N+s)!s!2^s}.$$

$$P_0(m \mid N, t) = \left(1 - \frac{1}{t}\right)^N P_0(m \mid N, t-1)$$
$$+ \binom{N}{1}\frac{1}{t}\left(1 - \frac{1}{t}\right)^{N-1} P_0(m \mid N-1, t-1)$$
$$+ \cdots + \binom{N}{m}\left(\frac{1}{t}\right)^m\left(1 - \frac{1}{t}\right)^{N-m} P_0(m \mid N-m, t-1)$$

$$(9.5) \qquad = t\binom{N}{m}\frac{1}{t^m}\left(1 - \frac{1}{t}\right)^{N-m} P_0(m-1 \mid N-m, t-1)$$
$$+ \binom{t}{2}\frac{N!}{m!m!(N-2m)!}\frac{1}{t^{2m}}\left(1 - \frac{2}{t}\right)^{N-2m} P_0(m-1 \mid N-2m, t-2)$$
$$+ \binom{t}{3}\frac{N!}{m!m!m!(N-3r)!}\frac{1}{t^{3m}}\left(1 - \frac{3}{t}\right)^{N-3m}$$
$$\cdot P_0(m-1 \mid N-3m, t-3) + \cdots.$$

10. Combinatorial formulae for chi-squared. In most statistical work the interesting values of χ^2 (or equivalently of S) are those greater than the expectation given the null hypothesis. If we are interested in $\Pr(S = M)$, where M is greater than the expectation, then $\Pr(S = 2M) + \Pr(S = 3M) + \cdots$ will be negligible and we get, from (5.2), to an adequate approximation (writing $\omega = \exp(2\pi i/M)$),

$$\Pr(S = M) = \Pr(S \equiv 0 \pmod M)$$

$$= \frac{N!}{t^N \cdot M} \text{ times the coefficient of } y^N \text{ in}$$

$$e^{ty} + \sum_{m=1}^{M-1} \left[\frac{1}{M} \sum_{r=0}^{M-1} \omega^{mr^2} \sum_{s=0}^{M-1} \omega^{-ms^2} e^{\omega^{2ms}y} \right]^t$$

if M is odd. The expression $\sum_r \omega^{mr^2}$ is the Gaussian sum and is equal to

$$(m/M)\sqrt{M}$$

or $i(m/M)\sqrt{M}$ according as $M \equiv 1$ or 3 (mod 4), where (m/M) is Legendre's symbol. The question arises whether the methods of Vinogradov [24] could be applied to the problem of the distribution of χ^2.

REFERENCES

[1] E. T. Copson, *The Asymptotic Expansion of a Function Defined by a Definite Integral of Contour Integral*, Admiralty Computing Service, London (1946).

[2] H. E. Daniels, "Saddlepoint approximations in statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 631–650.

[2A] G. Doetsch, *Theorie und Anwendung der Laplace-Transformation*, Springer, Berlin (1937); Dover, New York (1943).

[3] I. J. Good, *Probability and the Weighing of Evidence*, Griffin, London; Hafner Publishing Co., New York (1950).

[4] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, Vol. 40 (1953), pp. 237–264.

[5] I. J. Good, "A new finite series for Legendre polynomials," *Proc. Cambridge Philos. Soc.*, Vol. 51 (1955), pp. 385–388.

[6] E. Goursat, *Functions of a Complex Variable*, English translation by E. R. Hedrick and O. Dunkel, Ginn and Co., New York (1916), Vol. II, Part I.

[7] R. E. Greenwood and M. O. Glasgow, "Distribution of maximum and minimum frequencies in a sample drawn from a multinomial distribution," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 416–424.

[8] L. Guttman, "The test-retest reliability of qualitative data," *Psychometrika*, Vol. 11 (1946), pp. 81–95.

[9] G. H Hardy, *Ramanujan*, Cambridge University Press, Cambridge (1940).

[10) P. G. Hoel, *Introduction to Mathematical Statistics*, John Wiley Sons, New York; Chapman and Hall, London (1947).

[11] L. C. Hsu, "Approximations to a class of double integrals of functions of large numbers," *Amer. J. Math.*, Vol. 70 (1948), pp. 698–708.

[12] L. C. Hsu, "The asymptotic behaviour of a kind of multiple integrals involving a large parameter," *Quart. J. Math., Oxford Ser.* (2), Vol. 2 (1951), pp. 175–188.

[13] H. Jeffreys, *Theory of Probability*, Clarendon Press, Oxford (1939).

[14] D. G. KENDALL, "A summation formula for finite trigonometrical integrals," *Quart. J. Math., Oxford Ser.* (2), Vol. 13 (1942–3), pp. 172–184.

[15] M. G. KENDALL, "The derivation of multivariate sampling formulae from univariate formulae by symbolic operation," *Ann. Eugenics*, London, Vol. 10 (1941), pp. 392–402.

[16] M. G. KENDALL, *The Advanced Theory of Statistics*, Griffin, London, Vol. I (1945), Vol. II (1946).

[17] K. S. KRISHNAN, "On the equivalence of certain infinite series and the corresponding integrals," *J. Indian Math. Soc. (new series)*, Vol. 12 (1948), pp. 79–88.

[18] G. J. LIDSTONE, "Note on the general case of the Bayes-Laplace formula for inductive or *a posteriori* probabilities," *Trans. Fac. Actuaries*, Vol. 8 (1920), pp. 182–192.

[19] M. NASSIF, "On the behaviour of the function $f(z) = \sum_{n=0}^{\infty} e^{(2\pi)^{1/2} i n^2} \left(\frac{z^{2n}}{n!}\right)$," *Proc. London Math. Soc.* (2), Vol. 54 (1952), pp. 201–214.

[20] W. PERKS, "Some observations on inverse probability including a new indifference rule," *J. Inst. Actuar.*, Vol. 73 (1947), pp. 285–312.

[21] C. R. RAO AND I. M. CHAKRAVARTI, "Some small sample tests of significance for a Poisson distribution," *Biometrics*, Vol. 12 (1956), pp. 264–282.

[22] P. G. ROONEY, "Some remarks on Laplace's method," *Trans. Roy. Soc. Canada, Sect. III*, Vol. 47 (1953), pp. 29–34.

[23] S. R. TIMS, "Note on a paper by M. Nassif," *Proc. London Math. Soc.* (2), Vol. 54 (1952), pp. 215–218.

[24] I. M. VINOGRADOV, *The Method of Trigonometrical Sums in the Theory of Numbers*, translated into English by K. F. Roth and Anne Davenport, Interscience Publishers, London and New York (undated).

[25] W. A. WHITWORTH, *Choice and Chance*, 5th ed., Deighton and Bell, Cambridge (1901).

[26] S. S. WILKS, "The likelihood test of independence in contingency tables," *Ann. Math. Stat.*, Vol. 6 (1935), pp. 190–196.