MATHEMATICAL CENTRE TRACTS 31

W. MOLENAAR

APPROXIMATIONS TO THE POISSON, BINOMIAL AND HYPERGEOMETRIC DISTRIBUTION FUNCTIONS

MATHEMATISCH CENTRUM AMSTERDAM 1973

AMS (MOS) subject classification scheme (1970): 62F05, 62F25

1st. printing: 1970
2nd. printing: 1973

PREFACE

Approximations to well known discrete distribution functions are the subject of a large number of publications, but little has been done to provide guidelines for the choice of an approximation which is as accurate as possible, but simple enough for hand calculation. The present study tries to fill this gap, by a report on the asymptotic and numerical properties of the previously published approximations, and of some new ones.

Many thanks are due to Professors HEMELRIJK and VAN ZWET, for encouragement and criticism during the past two years. I am also grateful to Mr. VEHMEYER (for assistance in calculations and proofreading), Mrs. and Mr. HILLEBRAND (for typing and preparing illustrations), Mr. ZWARST and Mr. SUIKER (for reproduction). The possibility to use the Electrologica X8 computer of the Mathematisch Centrum was essential for the present investigation.

Amsterdam, 1970 — The author

PREFACE TO THE SECOND PRINTING

No attempt has been made to cover the literature of the four years that have elapsed since the first appearance of this monograph. Some minor slips have been corrected in this printing. One development not foreseen in 1970 is the large scale introduction of relatively cheap desk calculators and even pocket calculators that can do far more than their predecessors. My insistence on the desirability to avoid 'complicated' operations like third roots, natural logarithms and inverse sines (p.20) would not be so strong now.

The fact that my desk calculator can read 60-step programs, enabling exact evaluation of virtually all Poisson and binomial probabilities, has not seduced me to delate Chapters II and III: quite a few people have no immediate access to such beautiful gadgets.

Groningen, 1973 — The author

TABLE OF CONTENTS

SUMMARY

In mathematical statistics the Poisson, binomial and hypergeometric probability distributions are frequently used. In most applications it is desired to evaluate the (cumulative) distribution function for given values of the argument and the parameter(s). Computation of this function is usually not possible without an electronic computer. As the distribution functions studied depend on two, three and four variables respectively, the published tables cover only a limited range; moreover, accurate interpolation is cumbersome.

For this reason, approximations were proposed at an early stage of the development of probability theory and mathematical statistics. The normal distribution can be used for approximation to all three distributions. When a cumulative Poisson table or nomogram is available, one may also use a Poisson approximation to the binomial or hypergeometric distribution. For the latter, a binomial approximation can be used when binomial tables are available. These possibilities determine the structure of the present study, which is indicated in the following diagram.

Approximations to

		Poisson	Binomial	Hypergeometric
Introduction		II.1, p.27	III.1, p.70	IV.1, p.116
Numerical information		II.6, p.63	III.7, p.109	IV.5, p.147
General advice		II.6, p.64	III.7, p.110	IV.5, p.148
Approx. by				
	Normal	II.2-5, p.31	III.2-5, p.71	IV.2, p.119
	Poisson	-	III.6, p.105	IV.3, p.136
	Binomial	-	-	IV.4, p.141

The first three sections of the introductory Chapter I are essential for an understanding of Chapters II, III and IV.

In many situations the classical approximations are not accurate enough. Numerous publications are devoted to more refined approximations. Most

refinements, however, are too complicated to be attractive for hand calculation. In our computer era, they have only a limited importance, because a computer gives a quick and accurate answer by exact calculation (except in some extreme cases).

Simple approximations which are superior to the classical ones can be very useful. Frequently a statistician wants a quick evaluation of a tail probability. Hand calculation with a minimum of tables would enable him to obtain the answer without delay. Why should we force him to consult a rather special table in the library, or to wait for access to a computer?

The present study intends to obtain a maximum of accuracy with a minimum of computation. However, this problem has no optimal solution. There is no objective measure for the laboriousness of an approximation, as it depends on the nature of the given values of the parameters and argument, and on the availability of certain tables or computational facilities. Even a judgement on the accuracy contains subjective elements: the same approximation may produce large errors near the median and small errors near probabilities of .05. Moreover, even a generally bad approximation is usually very accurate for some special values of parameters and argument.

Our comparison of previously published approximations and our search for better ones relies on conclusions from asymptotic expansions of the errors. We made a numerical check of the conclusions for many (finite) parameter values. Generally speaking the agreement was good, but there are nearly always some exceptions. The simple advice given on pages 64, 110 and 148 could only be formulated at the cost of disregarding such exceptions. For obvious reasons, we have attached some extra importance to accurate approximation of probabilities near the customary significance levels or their complements.

A key result of our investigation is the advice to use

$$(1)\qquad \Phi(2\{N-1\}^{-\frac{1}{2}}\{(k+1)^{\frac{1}{2}}(N-n-r+k+1)^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(r-k)^{\frac{1}{2}}\})$$

as an approximation to the hypergeometric distribution function

$$(2) \qquad \sum_{j=0}^{k} \binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n} \qquad (0 \le k < n \le r \le \tfrac{1}{2}N).$$

In terms of a 2 × 2 table having fixed marginals, we may reformulate this advice as follows:

a) arrange rows and columns such that the upper left cell has smallest expected value;
b) take the square root of the product of the observed numbers on the main diagonal, each increased by one;
c) subtract a similar square root for the other diagonal, without adding ones;
d) multiply by $2(N-1)^{-\frac{1}{2}}$, where N denotes the grand total;
e) the standard normal distribution function Φ of the result is a good approximation to the probability of finding the observed number or less in the upper left cell.

Whereas (1) is especially accurate for probabilities of, say, less than .05 or more than .93, it can be replaced for the middle part of the distribution by

$$(3) \qquad \Phi(2N^{-\frac{1}{2}}\{(k+\tfrac{3}{4})^{\frac{1}{2}}(N-n-r+k+\tfrac{3}{4})^{\frac{1}{2}} - (n-k-\tfrac{1}{4})^{\frac{1}{2}}(r-k-\tfrac{1}{4})^{\frac{1}{2}}\}).$$

For $r < \frac{1}{2}N$ the result (3) is asymptotically twice as accurate as the classical normal and χ^2 approximations. In an extensive numerical investigation, also including the symmetric case $r = \frac{1}{2}N$, it was found to be frequently much better and rarely much worse. For probabilities below .05 and above .93 it is usually true that (1) is still better; it has the additional advantage of containing only three simple square roots of integers.

By an obvious limiting process, one obtains

$$(4) \qquad \Phi(2(k+1)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k)^{\frac{1}{2}}p^{\frac{1}{2}}) \qquad \text{(tails)}$$

$$(5) \qquad \Phi(2(k+\tfrac{3}{4})^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k-\tfrac{1}{4})^{\frac{1}{2}}p^{\frac{1}{2}}) \qquad \text{(middle part)}$$

as approximations to the probability of k or less successes in a binomial distribution with parameters n and p. The approximation (4) is twenty years old (FREEMAN & TUKEY, 1950), but its superiority over the classical $\Phi((k+\frac{1}{2}-np)(npq)^{-\frac{1}{2}})$ seems not to have been generally noticed. For p close to $\frac{1}{2}$, the approximations (4) and (5) remain accurate, but here it is still better to use $\Phi(2(k+5/8)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k-3/8)^{\frac{1}{2}}p^{\frac{1}{2}})$ for the middle part and (5) for tails.

For the probability of at most k events in a Poisson distribution with expectation λ, one similarly obtains

(6) $\Phi(2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$ (tails)

(7) $\Phi(2(k+\frac{3}{4})^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$ (middle part).

These normal approximations have been published, but not generally accepted, although they are rather more accurate than the classical $\Phi((k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}})$, and not more difficult. In Fig. 1 some graphs are given for the exact Poisson distribution function and the three just mentioned approximations. The size of the illustration makes an accurate reproduction of the actual errors difficult, but it is certainly visible that the classical normal approximation is inferior to (7) in the middle and to (6) in the tails of the distribution.

Although roughly twice as accurate as the classical approximations, the square root approximations (1), (3), (4), (5), (6), (7) may sometimes not be accurate enough. One may then use a more accurate, but also more cumbersome, normal approximation. Many approximations of this kind have been previously published. We have not only compared them, but we have also made a systematic study of improved forms of simple normal approximations. In most cases we could derive new approximations that are superior to the known ones in the sense of providing more accuracy for the same amount of calculation. The best choice from the more accurate approximations is included in our recommendations on pages 64, 110 and 148. We repeat that the evaluation of the accuracy and laboriousness of an

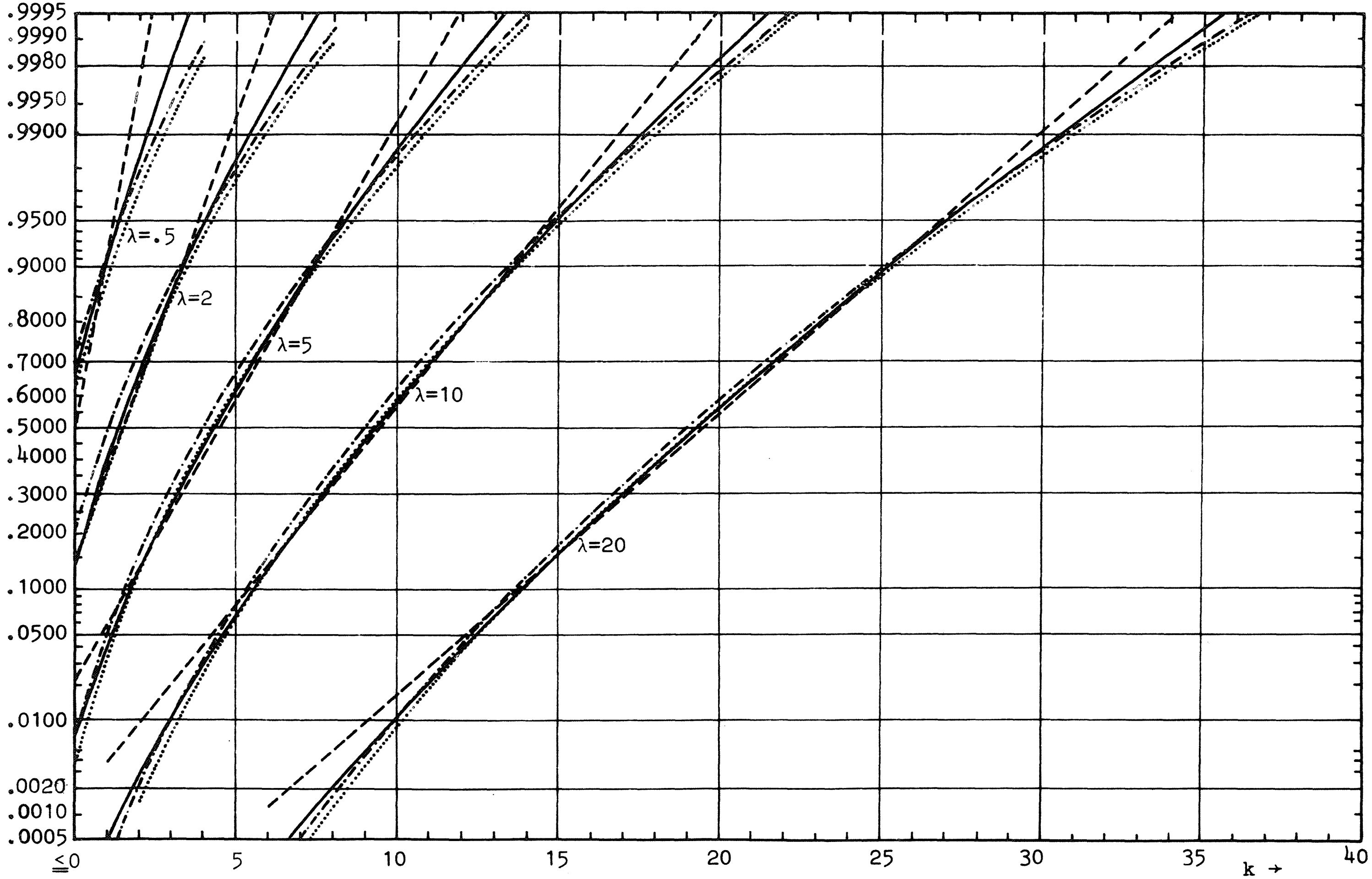

Fig. 1. Exact Poisson distribution function (———) and approximations $\Phi((k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}})$ (-------), $\Phi(2(k+\frac{3}{4})^{\frac{1}{2}}-2\lambda^{\frac{1}{2}})$ (.......) and $\Phi(2(k+1)^{\frac{1}{2}}-2\lambda^{\frac{1}{2}})$ (._._._.), for λ = .5, 2, 5, 10, 20

approximation inevitably contains a subjective element.

In the recommendations one also finds improved Poisson approximations to the binomial and hypergeometric distribution, and binomial approximations to the hypergeometric distribution. Their success can be explained by the fact that the parameters of the approximation contain not only the parameters of the unknown distribution function, but also its argument, which we shall always denote by k. For a simple Poisson approximation to the binomial probability of k or less successes, one should evaluate the analogous Poisson probability with parameter not np but rather

(8) $(2n-k)p/(2-p)$ (BOLSHEV, 1961);

or, still better, use

(9) $\{(12-2p)n-7k$[illegible]$p/\{(12-8p)n-k+k/n\}$.

Similarly the parameter of the Poisson approximation to the hypergeometric distribution should not be nrN^{-1} but rather

(10) $\frac{1}{2}(2n-k)(2r-k)/(2N-n-r+1)$,

again dependent on k. For a binomial approximation to the hypergeometric distribution, the success probability should not be rN^{-1} but rather

(11) $(2r-k)/(2N-n+1)$ (WISE, 1954).

For more detailed information we refer to sections III.6, IV.3 and IV.4, where the idea of a different approximating distribution for different values of the argument has been exploited in order to obtain small errors with a limited amount of computation.

Quantiles of the distributions, or confidence bounds for an unknown parameter, will not be explicitly considered. One could invert the results of sections II. 3d, III. 3d and IV. 2d. See also ANDERSON & BURSTEIN (1967,1968) and MOLENAAR (1969a).

CHAPTER I : GENERAL REMARKS

1. INTRODUCTION

This study tries to find simple and accurate approximations to the Poisson distribution function

(1.1) $$F_\lambda(k) = \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! \qquad (\lambda > 0;\ \text{integer } k \geq 0);$$

the binomial distribution function

(1.2) $$G_{n,p}(k) = \sum_{j=0}^{k} \binom{n}{j} p^j q^{n-j} \qquad (0 < p < 1;\ q = 1-p;\ \text{integer } n > 0,\ \text{integer } k,\ 0 \leq k \leq n);$$

and the hypergeometric distribution function

(1.3) $$H_{n,r,N}(k) = \sum_{j=0}^{k} \binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n} \qquad (\text{integer } k,n,r,N,\ 0 \leq k \leq n,\ 0 < n \leq r \leq \tfrac{1}{2}N).$$

The present Chapter gives a general introduction, some notations, an outline of the methods for comparing approximations, an explanation of what is meant by "simple and accurate", and some references. In Chapter II normal approximations to the Poisson distribution are considered in some detail: this simple situation with only one parameter offers a good opportunity for explaining the methods of analysis. The exposition of normal and Poisson approximations to the binomial distribution (Chapter III) proceeds along similar lines, and is therefore more condensed. The final Chapter IV is devoted to normal, Poisson and binomial approximations to the hypergeometric distribution.

Two main reasons can be given for a renewed study of this very classical subject of approximations to well-known discrete distributions. The first is the need of a comparison, as regards computational labour and accuracy, of the numerous approximations hitherto published. The second is the search for new approximations, or modified forms of existing approximations, which provide more accurate results without increasing the amount of computation.

This study is restricted to approximations to the distribution functions (1.1), (1.2) and (1.3). We have not included approximations to individual terms, to quantiles or to confidence limits for an unknown parameter. If we write, for the Poisson case, $F_\lambda(k) = P$, our goal is to find P from k and λ. The quantile problem, used e.g. in the determination of critical values for the testing of a hypothesis on λ, would be to find k for given P and λ. The confidence problem is essentially to find λ from k and P. In this sense a good approximation to P always leads to good approximations for quantiles or confidence bounds, but their calculation may be difficult when approximation to P is simple, or vice versa.

We have also excluded approximations combining an estimate of the first term in the tail with an estimate of the ratio of tail and first term (see e.g. BAHADUR (1960) for the binomial case) and saddle-point type approximations, usually too complicated to be qualified as "simple" approximations. Both excluded classes tend to be somewhat inaccurate whenever the tail consists of many terms; when this is not the case explicit summation of the terms may be a better solution.

The comparison of approximations as regards accuracy and simplicity inevitably introduces some subjective elements. We discuss our concept of "simple and accurate" in section 3, but let us state right now that there are no completely objective and uniformly valid criteria for the amount of work involved in the use of an approximation. Potential users may or may not have immediate access to a slide rule, a desk calculator or an electronic computer. Tables of logarithms, square roots or inverse sines may or may not be available. Parameter values are sometimes integers or simple fractions, sometimes numbers in five decimal places. Some statisticians want a simple trick giving the answer in thirty seconds, others do not mind if the calculation takes five minutes.

The evaluation of "accuracy" also contains subjective elements. In many situations we shall meet one approximation which is very accurate for probabilities, say, between .2 and .8, whereas another approximation is poor in this middle region, but rather accurate for probabilities of less than .1 or more than .9. For hypothesis testing with $\alpha = .05$ the latter may be preferable, but for estimating the median one would prefer the first. If one tries to attach one measure of accuracy to the approximations for different arguments k but for one set of parameters, one may choose from some five or more measures which have been proposed by different authors, cf. section 4.

Our comparison of approximations thus bears some resemblance to a report of a Consumers' Organization on washing machines or motor cars: the potential users differ widely in available facilities, frequency and circumstances of use, and wishes as regards the performance. A Consumers' Union tries to give objective information about advantages and disadvantages of the various products on the market. The final choice is left to the potential buyer, but some guidance is given how he may get the best possible performance for the amount of money he is prepared to spend.

In our case we shall present asymptotic and numerical results about the errors of various approximations, and give advice which approximation(s) could best be selected for quick work, for better approximation or for very accurate work. This advice relies on our judgement about accuracy and laboriousness, which inevitably involves some subjective elements. It is hoped that the reader of this study will bear this in mind when he sometimes disagrees with our conclusions. Section 3 of this Chapter considers in more detail the principles which guided our choice. It is nearly always true that more accuracy means more computation, but we try to select the best approximation from a class of equally laborious ones.

2. NOTATIONS AND METHODS

Tables, figures, formulae etc. are consecutively numbered within each section: "Fig.4.1" or (4.4)" indicate the first figure or fourth formula respectively of section 4 of the Chapter in which the reference occurs. For reference to other Chapters we add "of Chapter x", or write e.g. II (4.4). When a section is divided into subsections (4a, 4b, etc.) the numbering does not start anew at the beginning of new subsections. "Section 2" refers to section 2 of the same Chapter, and "section IV.3" to section 3 of

Chapter IV.

A list of symbols is included at the end of this study. Here we shall explain just a few notations, and define a few terms to which we give a special meaning. The notations F_λ, $G_{n,p}$ and $H_{n,r,N}$ for the Poisson, binomial and hypergeometric distribution function have been explained in (1.1), (1.2) and (1.3). We shall use

$$(2.1) \qquad \Phi(u) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{u} \exp(-\tfrac{1}{2}t^2)\, dt$$

and

$$(2.2) \qquad \varphi(t) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}t^2)$$

for the standard normal distribution function and density function respectively. Random variables will be distinguished from algebraic variables by underlining their symbols, e.g. we may put $P[\underline{x} = j] = e^{-\lambda}\lambda^j/j!$ if $\underline{x}$ has a Poisson distribution with expectation λ.

The word *tail* is always used for the minimum of the distribution function and its complement. When we say that a certain approximation overestimates left hand tails and underestimates right hand tails, it means that the exact distribution function is always smaller than its approximated value.

When P^* is an approximation to the value P of a distribution function, we shall sometimes use the *error* P^*-P, and sometimes the *relative tail error*, which is defined as

$$(2.3) \qquad \begin{cases} 100(P^*-P)/P & \text{if } P \leq .5, \\ 100\{(1-P^*) - (1-P)\}/(1-P) & \text{if } P > .5. \end{cases}$$

In the example P = .990, P^* = .995, the relative error in the distribution function is +.5 per cent., but one essentially uses $1-P^*$ = .005 as an approximation to the right hand tail 1-P = .010, which means a relative error of -50 per cent. For this reason we frequently use (2.3) as a measure of accuracy. For very small tails it usually becomes excessively large, but we shall never consider tails of less than .001, for reasons explained in section 3.

Any normal approximation to $F_\lambda(k)$, $G_{n,p}(k)$ or $H_{n,r,N}(k)$ can (and will) be written as $\Phi(u)$, where Φ is defined in (2.1), and u is a function of the argument k and the parameter(s) [λ, or n and p, or n, r and N] of the unknown distribution function. Such a function u, intended to be the argument of Φ, will be called a *normal deviate*, or briefly a *deviate*.

The well known order symbols of LANDAU-BACHMANN will be frequently used: $V(x) = O(W(x))$ $(x\to\infty)$ meaning that $|V(x)/W(x)|$ is bounded for sufficiently large x, and $V(x) = o(W(x))$ $(x\to\infty)$ meaning that $|V(x)/W(x)|$ converges to zero. When V or W is a function of more than one argument, it should be specified which variable tends to infinity, and in how far the boundedness or convergence is uniform in the other arguments. It would be cumbersome to repeat such a specification on each occasion, as it will usually remain unchanged as long as the approximating and approximated distribution are the same. If no explicit specification is given, it is therefore tacitly understood that a fixed set of assumptions is valid, which is given:

for normal approximations to the Poisson distribution in section II.2;
for normal approximations to the binomial distribution in section III.2;
for Poisson approximations to the binomial distribution in section III.6;
for normal approximations to the hypergeometric distr. in section IV.2a;
for Poisson approximations to the hypergeometric distr. in section IV.3;
for binomial approximations to the hypergeometric distr. in section IV.4.

Our main tool for comparing any two given approximations is the study of the asymptotic expansions of their errors, followed by a numerical verification of the properties deduced from the leading terms. We quote a simple example for normal approximations to the Poisson distribution function $F_\lambda(k)$. As explained in section II.2, there is a unique *exact normal deviate* ξ defined by $\Phi(\xi) = F_\lambda(k)$, but it is not explicitly known as a function of k and λ. The expressions $(k-\lambda)\lambda^{-\frac{1}{2}}$ and $(k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$ are known to differ little from ξ whenever λ is large and $|\xi|$ is not. Now it is found in Chapter II that for $\lambda \to \infty$ and bounded ξ one has

$$(2.4) \qquad \Phi((k-\lambda)\lambda^{-\frac{1}{2}}) - F_\lambda(k) = \phi(\xi)\,\lambda^{-\frac{1}{2}}\,(\xi^2 - 4)/6 + O(\lambda^{-1});$$

$$(2.5) \qquad \Phi((k+\tfrac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}) - F_\lambda(k) = \phi(\xi)\ \lambda^{-\frac{1}{2}}\ (\xi^2 - 1)/6 + O(\lambda^{-1}).$$

When λ is large enough to make the $O(\lambda^{-1})$ terms negligible compared to the leading terms, it follows that the former expression is more accurate whenever $|\xi^2 - 4| < |\xi^2 - 1|$, i.e. $\xi^2 > 2.5$. Now ξ has been defined by $\Phi(\xi) = F_\lambda(k)$, and $\Phi(-\sqrt{2.5}) = .057$. Thus we conclude for large values of λ that the continuity correction of $\frac{1}{2}$ had better be omitted whenever the desired probability is less than .057 or more than .943. Numerical investigation shows that the actual boundary values for $\lambda = 100$ are .060 and .946; for $\lambda = 4$ they are .073 and .958. In this case the asymptotic conclusion remains roughly valid for relatively small values of λ, although the numerical values are somewhat changed by the influence of terms of higher order.

In less simple cases the agreement between numerical results for finite parameter values, and conclusions from asymptotic expansions, is sometimes less satisfactory. It is a frequently encountered situation that the first term of an asymptotic series is much larger than the remaining terms together, but if the first two terms are included in some approximation, the third term, then leading term of the error, may be hardly larger than the fourth term unless the parameter tending to infinity is very large indeed.

Our main results on normal approximations are applications of a suitable continuity correction (either a constant or a function of some simple *normal deviate*) to the original random variable or its square root. For normal approximations to the Poisson distribution, for example, one gets much closer to the unknown value ξ if $v = (k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$ is replaced by $v^* = (k + (4-v^2)/6 - \lambda)\lambda^{-\frac{1}{2}}$. See also section 4 of Chapter II, where still better solutions of this type are given. A similar argument can be used to make simple approximations especially accurate near prescribed values of the unknown distribution function. This is worked out e.g. in section 3d of Chapter II.

We shall say that an approximation as occurring in (2.4) "has an error $O(\lambda^{-\frac{1}{2}})$". According to the formal definition of the O-symbol, the error is also $O(\lambda^{-\frac{1}{4}})$ or $O(1)$, but in such situations we shall use the sharpest generally possible exponent. There exist, however, nearly always exceptional cases for which a sharper exponent is possible. In the example (2.4) the error becomes "locally" $O(\lambda^{-1})$ for $\xi = \pm 2$. For normal approximations to the binomial distribution, we use $O(\sigma^{-1})$, where $\sigma^2 = npq$, even when the error is of a smaller order of magnitude for some special ξ if $p \neq q$, and for all ξ if $p = q$. These conventions avoid the repeated use of cumbersome statements on exceptions. In such situations the formulae will show for which special cases the error is of a smaller order of magnitude than formulated. The just mentioned conventions allow us to conclude that for large values of σ an approximation with error $O(\sigma^{-2})$ will be generally better than one with error $O(\sigma^{-1})$. In such a conclusion, exceptions as mentioned above are disregarded. Throughout this study, they are indeed exceptions, occurring only for special values of the unknown probability and/or of the parameters of the distribution function.

For Poisson or binomial approximations, the essential idea is to choose a parameter of the approximating distribution not only dependent on the parameter of the unknown distribution function, but also on its argument. The simplest example is the probability $G_{n,p}(k)$ of at most k successes in a binomial distribution with parameters n and p, which is roughly equal to the corresponding probability in a Poisson distribution with expectation np: the agreement becomes much better if np is replaced by e.g. $(2n-k)p/(2-p)$, a Poisson parameter depending also on the argument k.

The numerical calculations were all carried out on the Electrologica X8 computer of the Mathematisch Centrum, for which the author wrote the programmes in Algol 60.

In the early stages of the research, the long and tedious algebraic computations were carried out by hand, and checked by independent recomputation by Mr. VEHMEYER, whose assistance was very valuable. Later on, a different method was used for the algebraic operations with series expansions. They are usually rather straightforward from a mathematical point of view, but they involve three or sometimes five terms, and the terms of higher order are complicated expressions.

The use of a Formula Manipulation system, constructed by VAN DE RIET (1968), made it possible to leave most of the heavy arithmetic to the computer.

As Formula Manipulation is relatively new, it may be worthwhile to give a simple example of a programme. In section III.2 we shall meet an expansion (for $\sigma \to \infty$), which was given by PEIZER & PRATT (1968) in the form

$$\xi = u + \sigma^{-1}(q-p)(-u^2+1)/6 + \sigma^{-2}\{(1-4pq)(5u^3-2u)/72+pq(u^3-u)/12\} +$$

$$+ \sigma^{-3}\{(q-p)^3(-249u^4+79u^2+128)/6480+(q-p)pq(-31u^4+16u^2+27)/360\} + \quad (2.6)$$

$$+ O(\sigma^{-4}).$$

It was desired to invert this into an expansion for u with coefficients depending on ξ. To this end we fed into the computer the tape containing VAN DE RIET's formal system, and a second tape, containing (after an introductory comment and some declarations) the text shown in the first half of Table 2.1.

In the system, "TPS(y, a_0, a_1, ..., a_j)" indicates the Truncated Power Series $a_0 + a_1y + \dots + a_jy^j$. We use "isd" for σ^{-1} (inverse standard deviation) and "ISD" for the TPS $0.\sigma^0 + 1.\sigma^{-1} + 0.\sigma^{-2} + 0.\sigma^{-3}$. The desired expansion for u is: $\sum \sum x_{ij} \xi^j\sigma^{-i}$. We introduce it as shown in the programme, using our prior knowledge that several coefficients are zero, and indicating ξ by "ks". The fifth and sixth line of the programme describe the substitution of u and q into (2.6). The result is a TPS in ξ and σ^{-1} which is printed out. The unknowns are found by equating the coefficients of powers of "ks" in the coefficients of powers of "isd" to zero: the statement "SOL LIN EQ (...)" instructs the computer to solve the seven unknowns successively from the seven equations k10 = 0, k20 = 0, etc. The second half of Table 2.1 gives the output of the programme.

3. WHAT IS "SIMPLE AND ACCURATE"?

The purpose of this study is to find simple but accurate approximations. Let us sketch for what type of potential user our study of approximations is mainly intended to be useful. He wants to evaluate the Poisson, binomial or hypergeometric distribution function "without leaving his seat":

TABLE 2.1. Input and output of a Formula Manipulation programme explained in the text. The underlined number was actually printed as a real number in the output, because of the precision chosen for the programme.

```
PR STRING(results fbwml);
ISD:=TPS(isd,0,1,0,0); ISD2:=TPS(isd,0,0,1,0); ISD3:=TPS(isd,0,0,0,1);
u:=TPS(isd,TPS(ks,0,1,0,0,0),TPS(ks,x10,0,x12,0,0),TPS(ks,0,x21,0,x23,0),TPS(ks,x30,0,x32,0,x34));
u2:=u×u; u3:=u2×u; u4:=u3×u; q:=1-p;
ksi:= u + (q-p)×(1-u2)/6×ISD + (1-4×p×q)×(5×u3-2×u)/72×ISD2 + p×q×(u3-u)/12×ISD2
          + (q-p)↑3×(-249×u4+79×u2+128)/6480×ISD3 + (q-p)×p×q×(-31×u4+16×u2+27)/360×ISD3;
OUTPUT R(ksi:=ksi);
COEFF(ksi,k0,k1,k2,k3);
COEFF(k1,k10,k11,k12,k13,k14); COEFF(k2,k20,k21,k22,k23,k24); COEFF(k3,k30,k31,k32,k33,k34);
SOL LIN EQ(7,x10,x12,x21,x23,x30,x32,x34,
            k10,k12,k21,k23,k30,k32,k34);
END;
```

```
results fbwml
                 ksi:= 0+1×ks+0×ks↑2+0×ks↑3+0×ks↑4+O(ks↑5)+(x10-1/3×p+1/6+0×ks+(x12+1/3×p
-1/6)×ks↑2+0×ks↑3+0×ks↑4+O(ks↑5))×isd+(0+(2/3×x10×p-1/36×p↑2-1/3×x10+x21+1/36×p-1/36)×ks
+0×ks↑2+(2/3×x12×p+7/36×p↑2-1/3×x12+x23-7/36×p+5/72)×ks↑3+0×ks↑4+O(ks↑5))×isd↑2+(1/3×x10
↑2×p-1/36×x10×p↑2-13/1620×p↑3-1/6×x10↑2+1/36×x10×p+13/1080×p↑2-1/36×x10+x30-47/1080×p+
8/405+0×ks+(2/3×x10×x12×p+7/12×x10×p↑2-1/36×x12×p↑2-7/810×p↑3-1/3×x10×x12-7/12×x10×p+1
/36×x12×p+2/3×x21×p+7/540×p↑2+5/24×x10-1/36×x12-1/3×x21+x32-31/1080×p+79/6480)×ks↑2+0×
ks↑3+(1/3×x12↑2×p+7/12×x12×p↑2+73/540×p↑3-1/6×x12↑2-7/12×x12×p+2/3×x23×p-73/360×p↑2+5/
24×x12-1/3×x23+x34+13/90×p-83/2160)×ks↑4+O(ks↑5))×isd↑3+O(isd↑4);
                   x10:= 1/3×p-1/6;
                   x12:= -1/3×p+1/6;
                   x21:= -7/36×p↑2+7/36×p-1/36;
                   x23:= 1/36×p↑2-1/36×p-1/72;
                   x30:= -8/405×p↑3+4/135×p↑2+4/135×p-8/405;
                   x32:= 7/810×p↑3-7/540×p↑2-7/540×p+7/810;
                   x34:= 1/270×p↑3-1/180×p↑2-1/180×p+1/270;
READY    LINE NUMBER = 14        EXECUTION TIME = 329 SEC
```

he does not go to the library to consult some very special table, or to some terminal where he can put his question to a high-speed electronic computer. In such situations hand calculation with a minimum of tables should suffice. We do suppose that our potential user has a table of the standard normal distribution function(this seems reasonable)and preferably a table of square roots. If he uses a Poisson or binomial approximation to a binomial or hypergeometric probability, he must also have a table or nomogram of the Poisson distribution function (cf. section II.1b) or the binomial distribution function (cf. section III.1). However, whenever possible we want to avoid the use of third roots, natural logarithms, or inverse sines. In our opinion approximations as given e.g. by BORGES (1970), PEIZER & PRATT (1968) or BOLSHEV (1961), involving the use of some very special function, are unattractive; the tables of these functions are evidently less bulky than, say, complete binomial tables, but our potential user, who wants his answer at once, will seldom have immediate access to such specialized tables. We do mention such a complicated approximation, though, whenever it is more accurate than less laborious approximations.

When immediate access to a high-speed computer is available, it will be nearly always possible to evaluate the distribution function by direct summation in (1.1), (1.2) or (1.3), obtaining each term from the preceding one. When the argument k is very large, or when the limited precision of the computer causes too large errors, one of our "very accurate" approximations may be useful for computer programming.

In the following situations it is essential that the approximation to be used is accurate at or near the values α and/or $1-\alpha$ of the distribution function:

a) in hypothesis testing (determination of critical values, or probabilities of exceedance for observed values), with a level of significance of α (one-sided case) or 2α (two-sided case);

b) for the determination of confidence bounds for an unknown parameter, with confidence coefficient $1 - \alpha$ (one-sided case) or $1 - 2\alpha$ (two-sided case).

Thus accurate approximation to tails between, say, .005 and .05, is of special importance in most statistical applications. As was stated in section 2, the word *tail* always denotes the minimum of P and 1-P, where P is the unknown value of the distribution function. A not too bad performance in the middle part of the distribution comes as a next desideratum. Accurate approximation to tails of less than .001 is rarely essential, and for that reason all numerical investigations in the present study are confined to values of the distribution function between .001 and .999.

4. SOME REFERENCES

There are many publications on approximations to Poisson, binomial and hypergeometric distribution functions. References are collected at the end of this book, in a list which usually indicates in which section the approximation is discussed. Undoubtedly some references will inadvertently have been missed: a really complete survey of all existing approximations is a heavy task.

With some exaggeration one might say that most publications derive one new approximation and give two or ten numerical examples, in which the new idea gives better results than two or three previously published formulae. It may be worse in other situations, or more cumbersome for calculations, but that is not always made clear. It seems to be a law of nature that most people start to look for a simple approximation, find one that is not very accurate, and drift off towards other formulae which become more and more accurate, but also more and more cumbersome.

In such a situation there is certainly a place for comparative studies, but only very few have been published. The purpose of this section is to mention some of them, together with some other general papers on our subject. References to other publications will be made in subsequent Chapters.

The work of RAFF (1955, 1956) on normal and Poisson approximations to the binomial distribution is based on the error criterion "largest error which can arise in estimating any sum of consecutive binomial terms with the specified parameters", i.e.

$$(4.1) \qquad \max_{0 \leq i \leq j \leq n} \left| G_{n,p}(j) - G_{n,p}(i) - A(j) + A(i) \right| ,$$

TABLE 4.1

Binomial distribution function for $n = 40$, $p = .3$, errors of normal approximations (4.2) and (4.3) and error criteria (4.1), (4.4) and (4.5).

k	$G_{n,p}(k)$	error of (4.2)	error of (4.3)
4	.0026	+.0023	-.0002
5	.0086	+.0038	-.0000
6	.0238	+.0051	+.0012
7	.0553	+.0050	+.0041
8	.1110	+.0026	+.0086
9	.1959	-.0017	+.0137
10	.3087	-.0064	+.0175
11	.4406	-.0091	+.0185
12	.5772	-.0087	+.0164
13	.7032	-.0056	+.0120
14	.8074	-.0016	+.0072
15	.8849	+.0015	+.0032
16	.9367	+.0031	+.0006
17	.9680	+.0031	-.0005
18	.9852	+.0023	-.0008
19	.9937	+.0014	-.0006
20	.9976	+.0007	-.0004
RAFF's criterion (4.1)		.0142	.0193
largest absol. error (4.4)		.0091	.0185
sum of absol. errors (4.5)		.0663	.1062
as (4.5) but indiv. terms		.0346	.0391

where A(k) denotes some approximation to the binomial distribution function $G_{n,p}(k)$. The two papers give numerical values of these errors (4.1) for several parameter pairs (n,p), for six approximations (1956) and for nine (1955). RAFF's conclusions coincide roughly with ours (good results for arcsin and CAMP-PAULSON approximations).

It should be observed that criteria like (4.1) work in favour of approximations which are accurate in the middle part of the distribution, where absolute errors tend to be large. As an illustration Table 4.1 compares the classical normal approximation

(4.2) $\Phi((k+\frac{1}{2}-np)(npq)^{-\frac{1}{2}})$

to the square root approximation

(4.3) $\Phi(2\{(k+1)q\}^{\frac{1}{2}} - 2\{(n-k)p\}^{\frac{1}{2}})$,

for a binomial distribution with $n = 40$ and $p = .3$. Not only RAFF's criterion (4.1), but also the criteria

(4.4) $\max_{0 \leq j \leq n} |G_{n,p}(j) - A(j)|$ (HODGES & LECAM, 1960),

(4.5) $\sum_{j=0}^{n} |G_{n,p}(j) - A(j)|$

and the sum (4.5) with the distribution functions replaced by the individual terms of the probability distribution (PROHOROV, 1953; LECAM, 1960), indicate that (4.2) is considerably better that (4.3). However, for any tail less than .07 the square root deviate (4.3) is superior, and its use seems appropriate whenever accuracy near the customary significance levels is essential. According to the Table it gives e.g. for an exact binomial probability of .0238 the value .0250, whereas (4.2) gives .0289.

BLOM (1954) studies a class of transformations of binomial, negative binomial, Poisson and gamma variables. For a sketch of his methods we consider the binomial case, where we put $f = (k+1)/(n+1)$. The exact normal deviate ξ is defined by $G_{n,p}(k) = \Phi(\xi)$, for notations cf. (1.2) and (2.1). Now by the relations between binomial, beta, F- and FISHER's z-distribution, $G_{n,p}(k)$ equals $P[\underline{z} > \frac{1}{2} \log (p/q) - \frac{1}{2} \log (f/(1-f))]$, where $\underline{z}$ has 2k + 2 and 2n - 2k degrees of freedom. From the CORNISH-FISHER (1937) expansion for $\underline{z}$, BLOM obtains, for any function Ψ satisfying certain mild assumptions,

(4.6) $\Psi(p) = \Psi(f) + a_1(n+1)^{-\frac{1}{2}} + a_2(n+1)^{-1} + 0(n^{-3/2}),\ n \to \infty,$

where $a_1 = - \xi f^{\frac{1}{2}} (1-f)^{\frac{1}{2}} \Psi'(f)$, and a_2 is a similar function of f and ξ involving also Ψ''. A transformation Ψ for which $\Psi(f) + a_1(n+1)^{-\frac{1}{2}}$ is for large n a good approximation to $\Psi(p)$, can be found by putting $a_2 = 0$. This leads to a differential equation for Ψ, with solution

$$(4.7)\qquad \Psi(y) \equiv \int_{y_0}^{y} t^{-\eta}(1-t)^{-\eta}dt, \text{ where } \eta = \frac{2}{3}(1-\xi^{-2}).$$

The result is applicable in the confidence problem, which is the solution of p from $G_{n,p}(k) = \alpha$ for given n, k and α. This gives an upper bound for the success probability p, with confidence coefficient $1-\alpha$, when k successes in n independent experiments have been observed. The analogous lower bound satisfies $1 - G_{n,p}(k-1) = \alpha$, and one may use (4.6) with obvious changes. An approximation to p is found as the solution for x of $\Psi(x) = \Psi(f) + a_1(n+1)^{-\frac{1}{2}}$, and this is a close approximation if Ψ is determined by (4.7). BLOM finds for $|\xi| = 1$ that $\eta = 0$, i.e. $\Psi(y) \equiv y$, for $|\xi| = 2$ that $\eta = \frac{1}{2}$, i.e. $\Psi(y) \equiv 2 \arcsin y^{\frac{1}{2}}$, and for $|\xi| = \infty$ that $\eta = 2/3$.

For the quantile problem, i.e. the solution of k from $G_{n,p}(k) = \alpha$ for given n, p, α, BLOM uses a series $\Psi(f) = \Psi(p) + b_1(n+1)^{-\frac{1}{2}} + b_2(n+1)^{-1} + O(n^{-3/2})$, where b_2 vanishes if Ψ is again given by (4.7), but now with $\eta = (1+2\xi^{-2})/3$. This means for $|\xi| = 1$ that $\eta = 1$, i.e. $\Psi(y) \equiv \log(y/(1-y))$, for $|\xi| = 2$ that $\eta = \frac{1}{2}$, again $\Psi(y) \equiv 2 \arcsin y^{\frac{1}{2}}$, for $|\xi| = \infty$ that $\eta = 1/3$, a transformation studied later by BORGES (1970).

In the present study we want to evaluate the binomial distribution function for given n,p,k; the preceding formulae use the value ξ that is in our case unknown. BLOM observes that some a priori knowledge of ξ may be available, and that corresponds to our idea of making an approximation most accurate near preassigned probability values. However, our proposal of different continuity corrections for different probabilities may have some advantages over BLOM's suggestion of different transformations for different probabilities. Moreover, the beta transformation (4.7) only assumes a simple form when η is 1, $\frac{1}{2}$ or 0.

BLOM's paper concludes with a similar investigation of optimal transformations for Poisson, χ^2 and negative binomial distributions. We shall mention his two approximations to the Poisson distribution function in section II.4.

PEIZER & PRATT (1968) give one normal approximation, applicable after suitable substitutions to binomial, beta, F, t, negative binomial, Pascal, Poisson, gamma and χ^2 distributions. They expand the exact normal deviate ξ in terms of one general standardized variable, and compare the result to similar expansions for some other approximating deviates. They give graphs for some function of the error, and a detailed investigation of its asymptotic behaviour in various limiting situations.

Because of the many similarities between the present study and the work of PEIZER & PRATT, it may be useful to point out some essential differences. We search for simple approximations, which should be accurate near the customary significance levels, whereas PEIZER & PRATT, aiming mainly at accuracy near the median and for very small tails, obtain a very accurate and general formula, which is rather complicated for hand calculation. Suitable continuity corrections, a major tool in the present study, are hardly considered by PEIZER & PRATT. We try to give a more direct presentation of the errors of the main approximations. The present work contains also approximations to the hypergeometric distribution, and a section on Poisson approximations to the binomial. On the other hand, we do not give explicit results for the distributions related to the binomial and Poisson distributions.

Many results could be first derived for the hypergeometric case, results for the binomial and Poisson case following as limiting cases. We have chosen for a separate presentation beginning with the simplest case, in the hope that it would make the main ideas more accessible: the complications grow, with the number of parameters, from the Poisson via the binomial towards the hypergeometric distribution. The main technique of comparing asymptotic expansions is certainly a common feature of the two publications (it was independently used by the present author, after a suggestion by VAN ZWET, at the time of appearance of PEIZER & PRATT's paper).

At the time of appearance of GEBHARDT (1969), the present study was being printed. At the relevant places in Chapters II and III no references to this paper could be inserted, but we shall give some comments here. The paper gives numerical values of RAFF's error criterion (4.1) for nine approximations to the binomial distribution function, followed by a discussion and a recommendation. We have already stated that for our purpose the criterion (4.1) is a somewhat inadequate measure of accuracy, as it is determined by the accuracy in the middle part of the distribution, disregarding the properties of the approximation for tails of, say less than .1.

In many respects our asymptotic investigations confirm GEBHARDT's numerical results. He observes that the normal approximation III (3.1) with $c = (2-p)/3$, $d = 1/3$, $\delta = 0$ is superior to the classical one with $c = \frac{1}{2}$, $d = \delta = 0$. Our expansion III (3.3) shows that the former reduces the error to $O(\sigma^{-3})$ for the median ($\xi = 0$), which is favourable for RAFF's criterion (4.1), but leads to extra large errors in the tails. Similarly GEBHARDT's choice $\beta = 1/6$, $\gamma = 1/3$, $\delta = 1/3$ for the arcsine approximation III (3.14) gives a lower error near the median than RAFF's choice $\beta = \gamma = \delta = 0$. The inferiority of the normal GRAM-CHARLIER approximation III (4.19), u = BI, $c = \frac{1}{2}$, $d = \delta = 0$, compared to the BORGES and CAMP-PAULSON approximations, is observed by GEBHARDT, and can be predicted from our Table 4.1 in section III. 4d.

The limiting values of (4.1) for $n \to \infty$, $p \to 0$, $np \to \mu$ suggest that the BORGES and CAMP-PAULSON approximations have almost the same accuracy for $1.5 \leq \mu \leq 50$, whereas BORGES is better for small values of μ. Now the BORGES deviate tends to II (4.3) with $\beta = 0$, and the CAMP-PAULSON deviate to the WILSON-HILFERTY deviate II (4.8), with leading term of the error ($\mu \to \infty$) equal to $\phi(\xi)(\xi^3-6\xi)/(216\mu)$ and $\phi(\xi)(3\xi-\xi^3)/(108\mu)$ respectively. Numerical investigation shows that (4.1) is mainly determined by the error for $|\xi| \simeq .8$, and thus our result for $\mu \to \infty$ is in agreement with the values of (4.1) for GEBHARDT's largest value $\mu = 50$.

We disagree with GEBHARDT's recommendation of the Poisson Gram-Charlier approximation III (6.1), which he prefers to the Bolshev Poisson approximation $F_{\lambda_2}(k)$. Moreover, we do not share his preference of the BORGES approximation to the CAMP-PAULSON approximation. The divergence of opinion stems from the use of a different error criterion, and from a different view on the laboriousness of certain approximations.

Apart from the seven approximations already mentioned, GEBHARDT considers the Poisson approximations with parameters np and λ_9 (see section II. 6). Tables of the maximal error (4.1) are given for nine values of n ($3 \leq n \leq 1000$) and nine values of p ($.001 \leq p \leq .5$), and also for $n \to \infty$, $p \to 0$, $np \to \mu$.

CHAPTER II : NORMAL APPROXIMATIONS TO THE POISSON DISTRIBUTION

1. INTRODUCTION

1a. NOTATIONS AND SUMMARY

Throughout this Chapter, λ is a positive real number, and k and ν are non-negative integers. The random variable $\underline{x}$ has a Poisson distribution with expectation λ, and F_λ denotes its distribution function:

$$(1.1) \qquad F_\lambda(k) = P\left[\underline{x} \le k\right] = \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! \ .$$

Furthermore, $\underline{\chi}_\nu^2$ denotes a random variable with a chi-squared distribution with ν degrees of freedom, and density

$$(1.2) \qquad 2^{-\frac{1}{2}\nu}\ z^{\frac{1}{2}\nu-1}\ \exp(-\tfrac{1}{2}z)/\Gamma(\tfrac{1}{2}\nu) \text{ for } z > 0,$$

and $\underline{\gamma}_\nu$ has a gamma distribution with shape parameter ν, and density

$$(1.3) \qquad z^{\nu-1}\ e^{-z}/\Gamma(\nu) \text{ for } z > 0.$$

In the sequel we shall frequently use the well known relation

$$(1.4) \qquad \begin{aligned} F_\lambda(k) &= \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! = \int_\lambda^\infty z^k\ e^{-z}/k!\ dz = \\ &= P\left[\underline{\gamma}_{k+1} > \lambda\right] = P\left[\underline{\chi}_{2k+2}^2 > 2\lambda\right], \end{aligned}$$

where the second equality is proved by repeated partial integration and the others are even more straightforward.

Section 1b briefly sketches some sources for exact values of the Poisson distribution $F_\lambda(k)$, for given λ and k. In section 1c the use of normal approximations is discussed in general terms; these approximations are divided into simple, better and very accurate ones. This classification is made according to the asymptotic order of the error. The three classes are discussed in sections 3, 4 and 5 respectively. Section 2 contains the

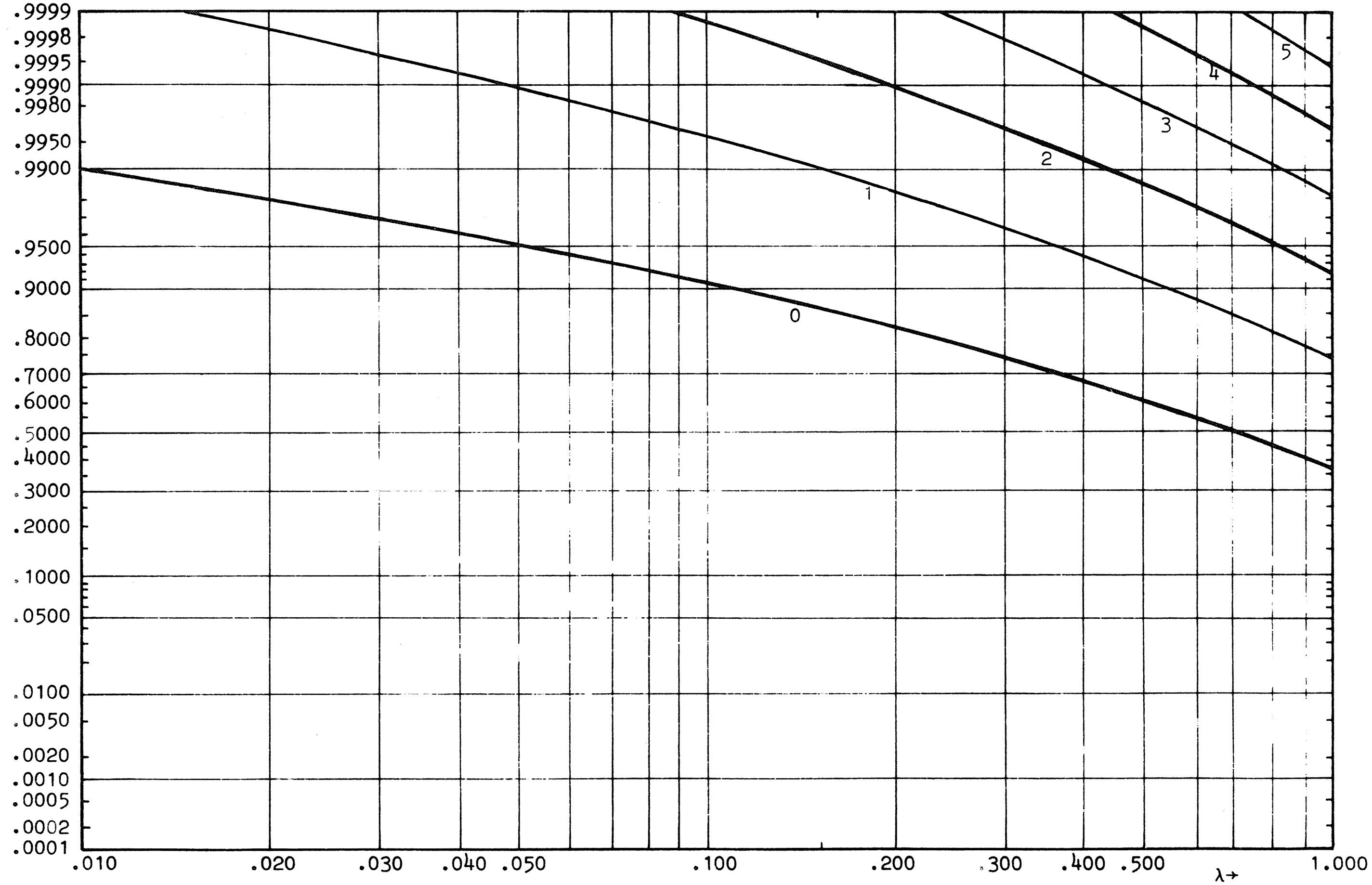

Fig. 1.1. Nomogram of the Poisson distribution function $F_\lambda(k)$ for $.01 < \lambda < 1$. Horizontal (logarithmic) scale for the parameter λ, vertical (normal probability) scale for $F_\lambda(k)$. Each curve belongs to one value of k.

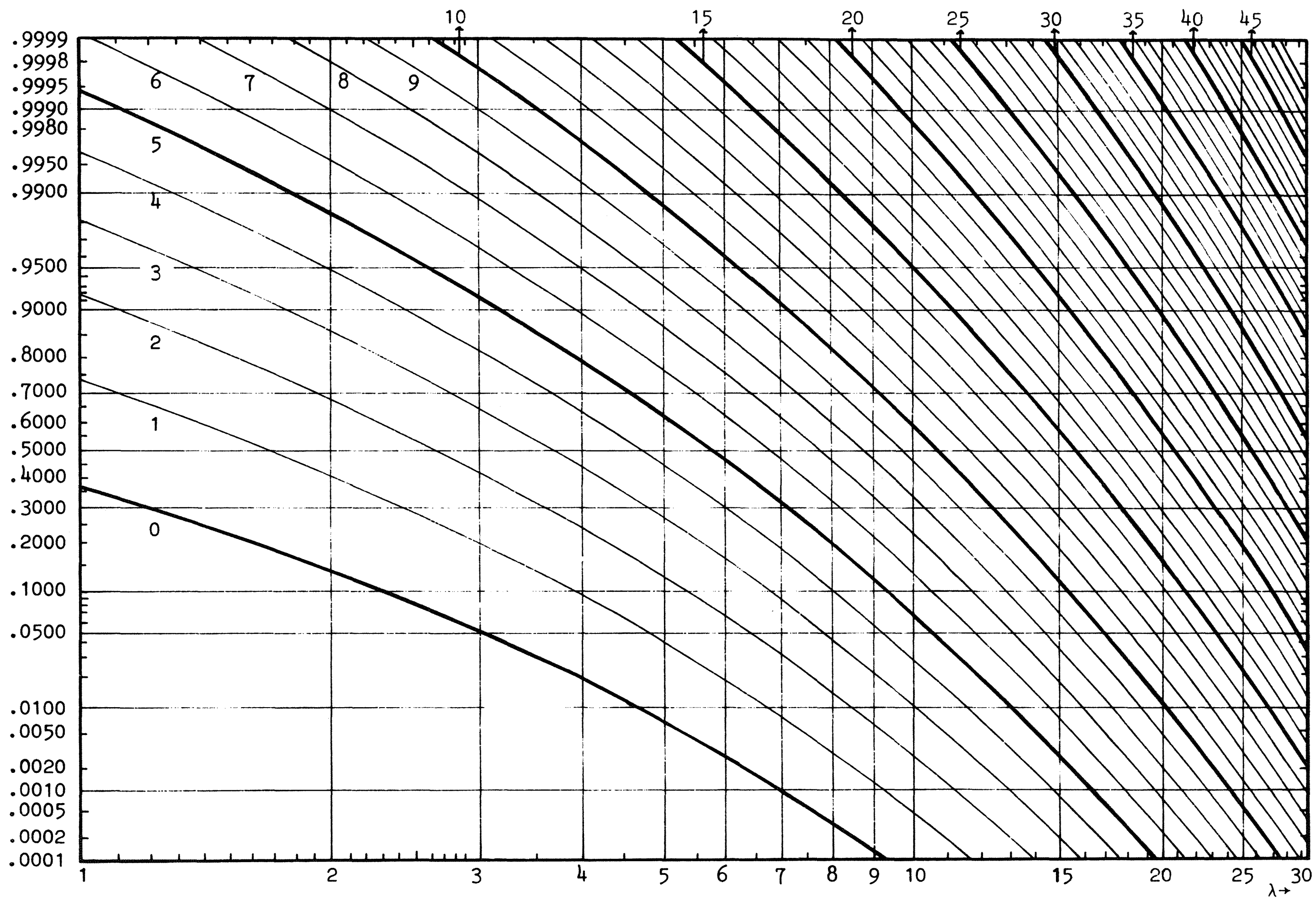

Fig. 1.2. Nomogram of the Poisson distribution function $F_\lambda(k)$ for $1 < \lambda < 30$. Horizontal (logarithmic) scale for the parameter λ, vertical (normal probability) scale for $F_\lambda(k)$. Each curve belongs to one value of k.

asymptotic results underlying the theoretical comparison of the various approximations, and section 6 gives numerical values of errors. An advice summarizing the main results is found at the end of the Chapter (Table 6.1, page 64).

1b. EXACT VALUES

Rough values of the Poisson distribution function $F_\lambda(k)$ can be read from a nomogram, as given in Figures 1.1 and 1.2. A slightly different form, with larger ranges for λ, k and probability, appeared already in CAMPBELL (1923). The tables of the VERENIGING VOOR STATISTIEK also contain a nomogram.

Special Poisson tables are e.g. MOLINA (1945), 6 decimals, λ =.001(.001).01(.01).3(.1)15(1)100, and GENERAL ELECTRIC (1962), 8 decimals, λ ranging from 10^{-7} to 205 with more intermediate values than MOLINA for $\lambda < 2$, but intervals of .5 for $5 < \lambda < 10$.

From (1.4) it is evident that one may also find $F_\lambda(k)$ from cumulative χ^2 tables such as PEARSON & HARTLEY (1954) Table 7, or from tables of the incomplete gamma function ratio (KHAMIS & RUDERT, 1965; PEARSON, 1922; HARTER, 1964), for the last two with a little extra calculation, as they give $P\left[\underline{Y}_\nu < z\nu^{\frac{1}{2}}\right]$ as a function of z and $\nu-1$.

For small values of λ, say $\lambda \leq .1$ or maybe $\lambda < .5$, it is easy for hand calculation to find first $e^{-\lambda}$, either from a table of the exponential function or from its series expansion, and to multiply it by the factor $(1+\lambda+\lambda^2/2!+\ldots+\lambda^k/k!)$. As e.g. $F_{.1}(2) = .9998$ and $F_{.5}(3) = .998$, this sum will rarely have more than a few terms.

1c. GENERAL REMARKS ON NORMAL APPROXIMATIONS

Our goal is to find functions u of k and λ such that $F_\lambda(k) \simeq \Phi(u)$; we recall that Φ and F_λ denote the standard normal and Poisson distribution function respectively. We shall use u or v as a general notation for such a *normal deviate*, and symbols like u_p, $u_{a,b}$ for special cases.

Normal approximations to F_λ have the general property of becoming more accurate when the parameter λ increases and the probability F_λ is kept constant. They are attractive when a cumulative normal table is directly available, but Poisson or similar tables are not. Normal tables are nearly always within immediate reach, because of their small size and general applicability. The equally small Poisson nomograms, like Figures 1.1 and 1.2, are less widespread, and when available may not be precise enough.

In section 2 series expansions will be considered for the exact solution ξ of the transcendental equation $F_\lambda(k) = \Phi(\xi)$. For simple approximations, i.e. with error $O(\lambda^{-\frac{1}{2}})$ for $\lambda \to \infty$ and constant probability, section 3 gives asymptotic and numerical results on errors. Section 4 contains a similar discussion for better normal approximations, with error $O(\lambda^{-1})$, and section 5 treats the very accurate ones with still smaller asymptotic errors. This classification according to the order of the error coincides almost completely with a classification according to the amount of computational work. However, within each class it is worthwhile to try to find the most accurate out of a group of equally laborious approximations. The evaluation of accuracy and computational labour inevitably contains a subjective element. We refer to section I.3 for a sketch of our principles regarding these aspects.

The main results on normal approximations to the Poisson distribution function are summarized in Table 6.1 at the end of this Chapter (page 64).

2. THE EXACT DEVIATE

Monotonicity considerations guarantee the existence of a unique *exact normal deviate* $\xi = \xi(k,\lambda)$ such that $\Phi(\xi) = F_\lambda(k)$. This section presents some asymptotic expansions for ξ, most of them well known, which form the basis of our investigations in sections 3, 4 and 5.

In all asymptotic expansions of this Chapter, it will be tacitly understood that $\lambda \to \infty$ and ξ is bounded. It is well known that this implies that $k \to \infty$ and that any deviate u for which $u - \xi = o(1)$ is also bounded. Because of the assumed boundedness, the order symbols can be considered to hold uniformly in ξ and k.

The restriction to bounded ξ, meaning that the values of the Poisson distribution function are bounded away from 0 and 1, is not serious for our purposes. We have already stated in Chapter I that accurate approximation of probabilities extremely close to 0 or 1 is rarely essential in current statistical practice. Values $|\xi| > 3.1$ (tail probabilities of less than about .001) have not been considered in our numerical investigations. The corresponding asymptotic results may safely be restricted to bounded deviates.

THEOREM 1.

The exact normal deviate ξ defined by $F_\lambda(k) = \Phi(\xi)$ satisfies the expansion

$$
\begin{aligned}
\xi = {} & u_\gamma + \\
& + (k+1)^{-\frac{1}{2}} (u_\gamma^2 - 1)/3 + \\
& + (k+1)^{-1} (7u_\gamma^3 - u_\gamma)/36 + \\
& + (k+1)^{-3/2} (219u_\gamma^4 - 14u_\gamma^2 - 13)/1620 + \\
& + O(k^{-2}),
\end{aligned}
\tag{2.1}
$$

where $u_\gamma = (k+1-\lambda)(k+1)^{-\frac{1}{2}}$.

PROOF. From (1.4) one finds $F_\lambda(k) = P[\underline{\gamma}_{k+1} > \lambda]$, and $\underline{\gamma}_{k+1}$ has r-th cumulant $(r-1)!(k+1)$. Thus $F_\lambda(k) = 1 - P[\underline{\gamma}^* \leq - u_\gamma]$, where $\underline{\gamma}^*$ denotes the standardized gamma variable $(\underline{\gamma}_{k+1} - E\underline{\gamma}_{k+1})/\sigma(\underline{\gamma}_{k+1})$. To the distribution of $\underline{\gamma}^*$ we apply the well known CORNISH-FISHER (1937) formulae (cf. HILL & DAVIS, 1968, formula (49) and Table 1), which are valid for the present case. For a somewhat different proof cf. RIORDAN (1949).

THEOREM 2.

The exact normal deviate ξ defined by $F_\lambda(k) = \Phi(\xi)$ satisfies the expansions

$$
\begin{aligned}
\xi = {} & u_p + \\
& + \lambda^{-\frac{1}{2}} (-u_p^2 + 1)/6 + \\
(2.2) \qquad & + \lambda^{-1} (5u_p^3 - 2u_p)/72 + \\
& + \lambda^{-3/2} (-249u_p^4 + 79u_p^2 + 128)/6480 + \\
& + O(\lambda^{-2}),
\end{aligned}
$$

where $u_p = (k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$, and

$$
\begin{aligned}
\xi = {} & u_s + \\
& + \lambda^{-\frac{1}{2}} (u_s^2 - 4)/12 + \\
(2.3) \qquad & + \lambda^{-1} (-u_s^3 + 10u_s)/72 + \\
& + \lambda^{-3/2} (21u_s^4 - 371u_s^2 - 52)/6480 + \\
& + O(\lambda^{-2}),
\end{aligned}
$$

where $u_s = 2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}$.

PROOF. From $k + 1 = \lambda + u_p\lambda^{\frac{1}{2}} + \frac{1}{2}$ follows

$$(2.4) \qquad u_\gamma = (u_p + \tfrac{1}{2}\lambda^{-\frac{1}{2}})(1 + u_p\lambda^{-\frac{1}{2}} + \tfrac{1}{2}\lambda^{-1})^{-\frac{1}{2}},$$

which can be expanded in powers of $\lambda^{-\frac{1}{2}}$ and substituted into (2.1). This leads to (2.2), and (2.3) follows by substitution of $u_p = u_s + (\frac{1}{4}u_s^2 - \frac{1}{2})\lambda^{-\frac{1}{2}}$ into (2.2). An alternative proof of (2.2) follows from direct application of the CORNISH-FISHER formulae to the Poisson variable, with correction for its lattice character (ESSEEN, 1945, p. 61). A third proof uses Stirling's formula for factorials for each Poisson term except the first terms (which give a negligible contribution), and combines them by the Euler summation formula. This third and rather tedious method will be used in Chapter IV for the hypergeometric distribution, where the other two methods break down. It may be observed that the well known (2.2) is a special case of expansion (5.1) of PEIZER & PRATT (1968, II).

3. SIMPLE APPROXIMATIONS

This section discusses the normal approximations which have error $O(\lambda^{-\frac{1}{2}})$ for $\lambda \to \infty$ and bounded deviate. The three main types of such approximations are considered in the first three subsections. Subsection 3d is applicable when special accuracy near preassigned probability levels is desired.

3a. SIMPLE POISSON TYPE

The Central Limit Theorem states that $(\underline{x}-\lambda)\lambda^{-\frac{1}{2}}$ is asymptotically normally distributed for $\lambda \to \infty$, where $\underline{x}$ has a Poisson distribution with $E\underline{x} = \sigma^2(\underline{x}) = \lambda$. Almost all authors apply a continuity correction of $\frac{1}{2}$: for the approximation of $F_\lambda(k) = P\left[\underline{x} \leq k\right]$ they use $\Phi(u_p)$, where $u_p = (k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$.

For hypothesis testing with a customary significance level α, however, we shall now show that this continuity correction produces extra large errors in the essential region of probabilities. From inversion of expansion (2.2) one obtains

$$(3.1)\qquad \begin{aligned} &(k+c-\lambda)\lambda^{-\frac{1}{2}} = \xi + \\ &+ \lambda^{-\frac{1}{2}}\,(\xi^2 + 6c-4)/6 + \\ &+ \lambda^{-1}\,(-\xi^3 - 2\xi)/72 + \\ &+ O(\lambda^{-3/2}), \end{aligned}$$

where the arbitrary constant c denotes the continuity correction. If one writes the right hand side of (3.1) as $\xi + R$, Taylor expansion leads to

$$(3.2)\qquad \begin{aligned} &\Phi((k+c-\lambda)\lambda^{-\frac{1}{2}}) - \Phi(\xi) = \\ &= R\Phi'(\xi) + \tfrac{1}{2}R^2\Phi''(\xi) + O(R^3) = \\ &= \phi(\xi)\,\{\lambda^{-\frac{1}{2}}\,(\xi^2 + 6c-4)/6 + \\ &+ \lambda^{-1}\,\left[-\xi^5 + (7-12c)\xi^3 + (-18+48c-36c^2)\xi\right]/72\} + \\ &+ O(\lambda^{-3/2}), \end{aligned}$$

where $\phi(\xi) = (2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}\xi^2)$ denotes the standard normal density function, and $\phi'(\xi) = -\xi\phi(\xi)$ has been used. According to (3.2) the leading term of the error $\Phi((k+c-\lambda)\lambda^{-\frac{1}{2}}) - F_\lambda(k)$ is proportional to $\xi^2 - 1$ for the usual choice $c = \frac{1}{2}$, but proportional to $\xi^2 - 4$ for $c = 0$. For large enough λ, the latter is thus optimal near $\xi = \pm 2$ (i.e. near the .023 and .977 fractiles) and the former near $\xi = \pm 1$ (.16 and .84 fractiles).

For tail probabilities *) of .057 or less, we have $\xi^2 > 2.5$, and consequently $|\xi^2 - 1| > |\xi^2 - 4|$. Then $(k-\lambda)\lambda^{-\frac{1}{2}}$ will lead to a smaller error than $(k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$, provided that λ is large enough to allow us to neglect terms of higher order in (3.2). Thus asymptotic behaviour indicates that the continuity correction of $\frac{1}{2}$ should only be used for probabilities between .057 and .943.

Numerical investigation shows that this region actually extends from .060 to .946 for $\lambda = 100$; from .063 to .949 for $\lambda = 25$; from .067 to .953 for $\lambda = 10$; from .073 to .958 for $\lambda = 4$. This discrepancy can be explained from the influence of the term order λ^{-1} in (3.2), which was neglected up to now. It works to the advantage of $c = 0$ for probabilities $\approx .057$, and to the advantage of $c = \frac{1}{2}$ for probabilities $\approx .943$, because the second term between curly brackets in (3.2) partially compensates the first one if $c = 0$ and $-2 < \xi < -1$, if $c = \frac{1}{2}$ and $1 < \xi < 2$. In the other two cases the two terms have the same sign.

By the same influence, the optimal accuracy of the deviate $(k-\lambda)\lambda^{-\frac{1}{2}}$ does not occur at probabilities of .023 and .977, as was just derived, but e.g. at .024 and .979 for $\lambda = 100$; at .028 and .981 for $\lambda = 10$. The optimum for $(k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$, claimed to occur at probabilities of .16 and .84, is actually found at .17 and .85 for $\lambda = 10$.

It follows from (3.2) that for $|\xi| > (4-6c)^{\frac{1}{2}}$ one has $R > 0$, and thus $\Phi((k+c-\lambda)\lambda^{-\frac{1}{2}}) > \Phi(\xi) = F_\lambda(k)$, whenever λ is large enough to make terms of higher order negligible. The use of the classical $(k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$ is expected

*) As stated in section I.2, *tail probability* means the smaller of the distribution function and its complement. As $F_\lambda(k) = \Phi(\xi)$ is the definition of ξ, we have $\xi < -\sqrt{2.5}$ if and only if $F_\lambda(k) < .057$ and $\xi > \sqrt{2.5}$ if and only if $1 - F_\lambda(k) < .057$.

to overestimate left hand tails of less than .16, and to underestimate right hand tails of less than .16. For tails exceeding .16 the reverse is expected to be true. For $(k-\lambda)\lambda^{-\frac{1}{2}}$ the preceding conclusion should hold with .16 replaced by .023. Numerical investigation confirms this, but the boundaries .16 and .023 are slightly different for moderate or small values of λ. Some examples have been given above.

It is evident that a continuity correction c varying with the value of ξ would improve the overall accuracy of the approximation. This idea is worked out in section 4. Special accuracy near some preassigned probability level can be achieved by a constant but suitable choice of c, as is shown in section 3d.

One could also try to improve the accuracy of the deviate $(k+c-\lambda)\lambda^{-\frac{1}{2}}$, for a fixed continuity correction c, by modifying the scaling factor $\lambda^{-\frac{1}{2}}$. It turns out that it would have to be replaced by $(\lambda + V\lambda^{\frac{1}{2}})^{-\frac{1}{2}}$, where V is some constant or simple function. This is very unattractive for hand calculation, so the idea was discarded.

3b. SIMPLE GAMMA TYPE

In the proof of Theorem 1 it was found that $u_\gamma = (k+1-\lambda)(k+1)^{-\frac{1}{2}}$ can also be used for normal approximation to $F_\lambda(k)$. Though a continuity correction seems out of place for the continuous gamma variable, we shall study the general *gamma type* deviate $(k+1+d-\lambda)(k+1)^{-\frac{1}{2}}$, where d is an arbitrary constant. From (2.1) or (3.1) follows after some calculation

$$(3.3)\qquad \begin{aligned} &(k+1+d-\lambda)(k+1)^{-\frac{1}{2}} = \xi + \\ &+ \lambda^{-\frac{1}{2}} (-\xi^2 + 1+3d)/3 + \\ &+ \lambda^{-1} \{7\xi^3 + (-13-18d)\xi\}/36 + \\ &+ O(\lambda^{-3/2}). \end{aligned}$$

Taylor expansion, like in (3.2), gives

$$\begin{aligned}(3.4)\qquad &\Phi((k+1+d-\lambda)(k+1)^{-\frac{1}{2}}) - \Phi(\xi) = \\ &= \phi(\xi)\ [\lambda^{-\frac{1}{2}}\ (-\xi^2 + 1+3d)/3 + \\ &+ \lambda^{-1}\ \{-2\xi^5 + (11+12d)\xi^3 + (-15-30d-18d^2)\xi\}/36] + \\ &+ 0(\lambda^{-3/2}).\end{aligned}$$

Let us suppose that λ is large enough to make terms of higher order than $\lambda^{-\frac{1}{2}}$ negligible. Comparing gamma type with given d to Poisson type with $c = (1-d)/2$, one finds that $\lambda^{-\frac{1}{2}}\ (-\xi^2 + 1+3d)/3 = -\ 2\ \lambda^{-\frac{1}{2}}\ (\xi^2 + 6c-4)/6$. According to (3.2) and (3.4) this means that for any given gamma type deviate there exists a corresponding Poisson type deviate for which the error is smaller by a factor 2, with opposite sign. This holds for all ξ, apart from the two values for which the leading term of the error vanishes. Whenever the Poisson parameter λ is large, it may thus be expected that gamma type deviates are relatively inaccurate. Therefore we shall condense our discussion of this type.

Still neglecting terms of higher order, one finds from (3.4) that the choice $d = 0$ produces optimal accuracy for $\xi = \pm 1$ (tails of .16), whereas $d = 1$ is optimal for $\xi = \pm 2$ (tails of .023).

Just as in section 3a one obtains asymptotic results on under- and overestimation, with all signs reversed.

Modification of the scaling factor $(k+1)^{-\frac{1}{2}}$ is unattractive for the same reason as mentioned for $\lambda^{-\frac{1}{2}}$ at the end of section 3a.

Numerical investigation revealed that the optimal accuracy for $d = 0$, claimed to be attained at probabilities of .16 and .84, actually was reached ad .14 and .82 for $\lambda = 10$. The optimal values of .023 and .977 for $d = 1$ were .020 and .976 for $\lambda = 10$. Just as in section 3a, the discrepancy can be explained if one considers the term of order λ^{-1} in (3.4).

Asymptotically gamma type $d = 1$ should have an error twice as large, and with opposite sign, when compared to Poisson type $c = 0$, and similarly for $d = 0$ and $c = \frac{1}{2}$. Of course the numerical error for finite values of λ cannot be expected to be exactly twice as large, but one may at least expect that Poisson type is more accurate. In the numerical investigation, exceptions to this were observed for $\lambda > 5$ only in the immediate vicinity of the optimal values (where errors are small anyhow), and for left tails of less than roughly .005 and $\lambda < 55$ (where gamma type d = 0 is more accurate than expected, by a compensation in terms of higher order). For $\lambda \leq 5$ exceptions were more numerous, but it remained true that Poisson type was usually more accurate.

If one standardizes not the gamma distribution, but the chi-squared distribution corresponding to it by (1.4), the result obviously does not change. However, one may also use FISHER's result that $(2\chi^2_\nu)^{\frac{1}{2}} - (2\nu-1)^{\frac{1}{2}}$ is for $\nu \to \infty$ asymptotically standard normal. As we have $F_\lambda(k) = P[\chi^2_{2k+2} > 2\lambda]$, this leads to $F_\lambda(k) \approx \Phi(2(k+\frac{3}{4})^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$. The accuracy of expressions of this type is investigated in section 3c. If one uses the still better WILSON-HILFERTY (1931) approximation to χ^2, one finds (4.8) with error $0(\lambda^{-1})$, see section 4a.

3c. SIMPLE SQUARE ROOT TYPE

In analysis of variance it would be useful to replace the Poisson variable $\underline{x}$, for which $\sigma^2(\underline{x}) = E\underline{x} = \lambda$, by some transformed variable $\Psi(\underline{x})$ with variance independent of its expectation. CURTISS (1943) has shown that this goal cannot be reached (except by the trivial transformation Ψ = constant). However, for any constant c the transformed variable $(\underline{x}+c)^{\frac{1}{2}}$ has the variance

$$(3.5) \qquad \sigma^2((\underline{x}+c)^{\frac{1}{2}}) = \tfrac{1}{4} + \lambda^{-1}\,(3-8c)/8 + \lambda^{-2}\,(17-52c+32c^2)/32 + 0(\lambda^{-3}),$$

which is at least asymptotically constant.

Moreover, it follows from a more general theorem of CURTISS (1943) that $(\underline{x}+c)^{\frac{1}{2}}$ is also asymptotically normal. Its expectation is

$$(3.6) \qquad E(\underline{x}+c)^{\frac{1}{2}} = (\lambda+c)^{\frac{1}{2}} - \lambda^{-\frac{1}{2}}/8 + 0(\lambda^{-3/2}).$$

ANSCOMBE (1948) mentions that the skewness coefficient of $(\underline{x}+c)^{\frac{1}{2}}$ is $(-\frac{1}{2})$ times the skewness coefficient of $\underline{x}$, and proposes the choice c = 3/8 for optimal variance stabilization, as this makes the term of order λ^{-1} vanish in

(3.5). BARTLETT (1936) uses $(\underline{x}+\frac{1}{2})^{\frac{1}{2}}$, FREEMAN & TUKEY (1950) use $\underline{x}^{\frac{1}{2}} + (\underline{x}+1)^{\frac{1}{2}}$ and LAUBSCHER (1960) uses $\underline{x}^{\frac{1}{2}} + (\underline{x}+\frac{3}{4})^{\frac{1}{2}}$, apparently because of empirically established good stabilization of the variance for moderate values of λ.

When a continuity correction $\frac{1}{2}$ is applied for the approximation to $F_{\lambda}(k) = P(\underline{x} \leq k)$, ANSCOMBE's variable $(\underline{x} + 3/8)^{\frac{1}{2}}$ leads to the standard normal deviate $2\ (k + 7/8)^{\frac{1}{2}} - 2(\lambda + 3/8)^{\frac{1}{2}} + 0(\lambda^{-\frac{1}{2}})$. From an approximation of FREEMAN & TUKEY (1950) for the binomial distribution follows the use of $u_s = 2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}$ for the Poisson case, cf. also BLOM (1954) formula (9.16). We shall study for any constants a and b the error of

$$u_{a,b} = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}. \tag{3.7}$$

From the expansion (3.1) one obtains after some calculation

$$\begin{aligned} u_{a,b} &= \xi + \\ &+ \lambda^{-\frac{1}{2}}\ (-\xi^2 + 12a-12b-8)/12 + \\ &+ \lambda^{-1}\ \{\xi^3 + (11-18a)\xi\}/36 + \\ &+ 0(\lambda^{-3/2}). \end{aligned} \tag{3.8}$$

In the notation $\delta = a-b$ it follows, just as in (3.2) and (3.4), that

$$\begin{aligned} &\Phi(u_{a,b}) - \Phi(\xi) = \\ &= \phi(\xi)\ [\lambda^{-\frac{1}{2}}\ (-\xi^2 + 12\delta-8)/12 + \\ &+ \lambda^{-1}\ \{-\xi^5 + (-8+24\delta)\xi^3 + (24-144b + 48\delta-144\delta^2)\xi\}/288] + \\ &+ 0(\lambda^{-3/2}). \end{aligned} \tag{3.9}$$

Now any choice of a and b with $\delta = a-b = 1$ makes the leading term of the error proportional to $\lambda^{-\frac{1}{2}}\ (4 - \xi^2)/12$, which means optimal accurary near $\xi = \pm 2$ (tails of .023), and any choice with $\delta = \frac{3}{4}$ gives an error asymptotically proportional to $\lambda^{-\frac{1}{2}}\ (1 - \xi^2)/12$, which means optimality near $\xi = \pm 1$ (tails of .16). Left hand tails of less than .16 (for $\delta = \frac{3}{4}$) or .023 (for $\delta = 1$) will be underestimated, and the corresponding right hand tails overestimated: the reverse should hold for larger tails.

For any Poisson type deviate with given c, a square root type deviate with $\delta = 1-\frac{1}{2}c$ has error $\lambda^{-\frac{1}{2}}\,(-\xi^2 + 12\delta-8)/12 = -\frac{1}{2}\,\lambda^{-\frac{1}{2}}\,(\xi^2 + 6c-4)/6$, and is thus asymptotically twice as accurate as the Poisson type, with opposite sign of error. In the same way, for any gamma type deviate there exists a square root type deviate which is asymptotically four times as accurate, with the same sign of error. These conclusions hold uniformly in ξ, apart from the two values of ξ for which the leading term of the error vanishes.

These asymptotic conclusions were checked in a numerical investigation, for which $b = 0$ was combined with $a = 1$ or $a = \frac{3}{4}$. The boundaries of .16 and .84 were actually .14 and .82 for $\lambda = 10$, and the boundaries of .023 and .977 were observed to be .020 and .975 for $\lambda = 100$, .013 and .970 for $\lambda = 10$. This is in accordance with the sign of the $O(\lambda^{-1})$ term in (3.9).

The numerical comparison of the three types revealed not exactly a 1 : 2 : 4 ratio for the errors, but square root type was at least more accurate than the others, with for any $\lambda > 5$ only a few exceptions near the boundary values (where errors are anyhow small). Even for λ between .5 and 5 the square root type was usually more accurate. Numerical values of its errors will be given in section 6.

The actual choice of a and b, under the restriction that $\delta = a-b$ equals 1 or $\frac{3}{4}$, only affects the $O(\lambda^{-1})$ term. For $\xi = \pm 2$ and $\delta = 1$, (3.9) becomes $\mp\,\phi(2)\,\lambda^{-1}\,(1+6b)/6 + O(\lambda^{-3/2})$, which indicates that $b = -\,1/6$, $a = 5/6$ might have some advantage over the customary choice $b = 0$, $a = 1$. Similarly, one obtains $\mp\,\phi(1)\,\lambda^{-1}\,(1+12b)/24 + O(\lambda^{-3/2})$ for $\xi = \pm 1$ and $\delta = \frac{3}{4}$, indicating that $b = -\,1/12$, $a = 2/3$ might be locally better than $b = 0$, $a = \frac{3}{4}$. However, due to the discreteness of the Poisson distribution, it is rather rare that a Poisson probability is extremely close to $\Phi(\pm 1)$ or $\Phi(\pm 2)$, and unless $|\xi|$ is very near to 1 or 2 one cannot expect a serious improvement from the more sophisticated choice of a and b. Numerical results confirm that it is not worthwhile to use $2(k + 5/6)^{\frac{1}{2}} - 2(\lambda - 1/6)^{\frac{1}{2}}$ instead of $2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}$, or $2(k + 2/3)^{\frac{1}{2}} - 2(\lambda - 1/12)^{\frac{1}{2}}$ instead of $2(k+\frac{3}{4})^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}$.

In section 3d, special but constant choices of a and b are investigated, which should provide optimal accuracy for preassigned values of the probability. In section 4d, a and b are made dependent on some simple deviate, which allows to reduce the error to $O(\lambda^{-1})$ or even $O(\lambda^{-3/2})$. The numerical errors for some simple constants a and b are presented in section 6.

The use of $\underline{x}^{\frac{1}{2}} + (\underline{x}+1)^{\frac{1}{2}}$ or $\underline{x}^{\frac{1}{2}} + (\underline{x}+\frac{3}{4})^{\frac{1}{2}}$ for variance stabilization suggests that $(k+a_1)^{\frac{1}{2}} + (k+a_2)^{\frac{1}{2}} - (\lambda+b_1)^{\frac{1}{2}} - (\lambda+b_2)^{\frac{1}{2}}$ could be used, instead of $u_{a,b}$ given by (3.7), as a normal deviate for approximation to $F_\lambda(k)$. However, it follows from (3.8) and (3.9) that it provides no essential improvement upon the choice $a_1 = a_2 = a$, $b_1 = b_2 = b$, investigated before, as the leading term of the error does not change. Numerical evidence confirms this conclusion.

3d. ACCURATE APPROXIMATION NEAR PREASSIGNED VALUES

Sometimes an approximation to cumulative Poisson probabilities is desired to be accurate near the prescribed values α and/or $1-\alpha$ of the distribution function, whereas it may be rather rough elsewhere. This will be the case when a two-sided hypothesis about λ is tested at significance level 2α, or a one-sided hypothesis at level α. We shall suppose that $\alpha < .5$.

Let ξ_α denote the standard normal upper α fractile, defined by $\Phi(\xi_\alpha) = 1-\alpha$. If one uses $u_{a,b}$ given by (3.7) with $\delta = a-b = (\xi_\alpha^2+8)/12$, the term of order $\lambda^{-\frac{1}{2}}$ in (3.9) vanishes for $|\xi| = \xi_\alpha$, i.e. for probabilities of α or $1-\alpha$. For the choice $a = (\xi_\alpha^2+11)/18$ and $b = (-\xi_\alpha^2-2)/36$, the term of order λ^{-1} vanishes too at $|\xi| = \xi_\alpha$; as before the error remains $O(\lambda^{-\frac{1}{2}})$ for all other ξ.

Similarly, one could use the Poisson type $(k+c-\lambda)\lambda^{-\frac{1}{2}}$ with $c = (4-\xi_\alpha^2)/6$, or the gamma type $(k+1+d-\lambda)(k+1)^{-\frac{1}{2}}$ with $d = (\xi_\alpha^2-1)/3$: both have a "local error" $O(\lambda^{-1})$ at $|\xi| = \xi_\alpha$ and an error $O(\lambda^{-\frac{1}{2}})$ for other ξ.

For the reader's convenience, values of the just mentioned functions of ξ_α are given in Table 3.1 for some customary values of α.

We are thus led to compare the accuracy of

$$(3.10) \qquad \Phi(2\{k + (\xi_\alpha^2+11)/18\}^{\frac{1}{2}} - 2\{\lambda - (\xi_\alpha^2+2)/36\}^{\frac{1}{2}}),$$

(3.11) $\Phi(\{k + (4-\xi_\alpha^2)/6 - \lambda\}\lambda^{-\frac{1}{2}})$,

(3.12) $\Phi(\{k + 1 + (\xi_\alpha^2-1)/3 - \lambda\}\{k+1\}^{-\frac{1}{2}})$,

(3.13) $\Phi(2\{k+1\}^{\frac{1}{2}} - 2\{\lambda + (4-\xi_\alpha^2)/12\}^{\frac{1}{2}})$,

where we have included (3.13) as a simple specimen of the square root type $u_{a,b}$ with $\delta = a-b = (\xi_\alpha^2+8)/12$. Table 3.2 gives the asymptotic errors, cal-

TABLE 3.1

Standard normal upper α fractiles ξ_α, and some functions of ξ_α used as corrections to simple deviates.

α	.1	.05	.025	.01	.005
ξ_α	1.281552	1.644854	1.959964	2.326348	2.575829
$\delta = (\xi_\alpha^2+8)/12$	+.8035	+.8921	+.9868	+1.1177	+1.2196
$a = (\xi_\alpha^2+11)/18$	+.7024	+.7614	+.8245	+ .9118	+ .9797
$-b = (\xi_\alpha^2+2)/36$	+.1012	+.1307	+.1623	+ .2059	+ .2399
$c = (4-\xi_\alpha^2)/6$	+.3929	+.2157	+.0264	- .2353	- .4391
$d = (\xi_\alpha^2-1)/3$	+.2141	+.5685	+.9472	+1.4706	+1.8783

TABLE 3.2

Leading terms of the errors for (3.10), (3.11), (3.12) and (3.13)

	$\lvert\xi\rvert \neq \xi_\alpha$	$\xi = \pm\xi_\alpha$
(3.10)	$\lambda^{-\frac{1}{2}}(\xi_\alpha^2 - \xi^2)\phi(\xi)/12$	$O(\lambda^{-3/2})$
(3.11)	$\lambda^{-\frac{1}{2}}(\xi^2 - \xi_\alpha^2)\phi(\xi)/6$	$\mp\lambda^{-1}(\xi_\alpha^3+2\xi_\alpha)\phi(\xi_\alpha)/72$
(3.12)	$\lambda^{-\frac{1}{2}}(\xi_\alpha^2 - \xi^2)\phi(\xi)/3$	$\pm\lambda^{-1}(\xi_\alpha^3-7\xi_\alpha)\phi(\xi_\alpha)/36$
(3.13)	$\lambda^{-\frac{1}{2}}(\xi_\alpha^2 - \xi^2)\phi(\xi)/12$	$\pm\lambda^{-1}(\xi_\alpha^3-7\xi_\alpha)\phi(\xi_\alpha)/36$

culated from (3.2), (3.4) and (3.9), for the special a, b, c, d considered here.

From the leading terms given in Table 3.2 we predict that (3.10) is superior throughout. Of the other three, which are all somewhat simpler, (3.11) is locally best at $|\xi| = \xi_\alpha$ unless $\xi_\alpha^3+2\xi_\alpha > |2\xi_\alpha^3-14\xi_\alpha|$, i.e. unless $2 < \xi_\alpha < 4$. By its definition, ξ_α is always non-negative. Now a given Poisson distribution hardly ever assumes values very close to α or $1-\alpha$, due to its discrete nature. And unless $\xi_\alpha^2 - \xi^2$ is very small, we may expect (3.13) to be better than (3.11) and (3.12).

The validity of these asymptotic conclusions was checked in a numerical investigation for α = .1, .05, .025, .01 and .005. In order to avoid the difficulties arising from the discrete character of the Poisson distribution, k was given some fixed value, and λ was selected to make the distribution function $F_\lambda(k)$ exactly equal to α or $1-\alpha$, and also to $.8\alpha$, $.95\alpha$, 1.05α, 1.2α or their complements. Absolute values of the errors were tabled, and it was also investigated for what values of the probability the errors were exactly zero.

The superiority of (3.10) was clearly confirmed for all $\lambda > 5$. The situation for the other three approximations was less clear. Especially for $\alpha \geq .025$, the error of (3.13) was zero at probability values which were even for $\lambda = 100$ (say) at some distance of α and $1-\alpha$. This phenomenon, due to the influence of terms of higher order, had some disturbing effect. However, it was usually true for probabilities of exactly α and $1-\alpha$, that (3.11) is better than (3.12) or (3.13) for $\alpha \geq .025$, and worse for $\alpha = .01$ and $\alpha = .005$ (which means indeed $2 < \xi_\alpha < 4$). For probabilities of $.8\alpha$, 1.2α or their complements, the 1 : 2 : 4 ratio of the absolute errors of (3.13), (3.11) and (3.12) was visible for (say) $\lambda > 150$, but for smaller λ there were more and more exceptions to the superiority of (3.13) over the other two.

Table 3.3 gives some examples of the errors of the four approximations, for $\alpha = .05$ and λ roughly 2, 10 and 100. The first line of the table means that the Poisson probability $F_{1.85}(4) = P[\underline{x} \leq 4 | \lambda = 1.85] = .9600$; substitution of $\xi_\alpha = 1.644854$, $k = 4$ and $\lambda = 1.85$ into (3.10) gives a value .9593 (error - .0007); similarly one finds .9592, .9519 and .9530 for (3.11), (3.12) and (3.13). The table does not give probabilities of .04, .05 and .06 for λ near 2: as e.g. $F_2(0)$ = .1353 and $F_3(0)$ = .0498, one only has left hand tails near α when k = 0, and the probability $F_\lambda(0) = e^{-\lambda}$ is simple enough for direct calculation.

As the four approximations do not differ much in computational labour, whereas (3.10) is considerably more accurate, we propose to use it for all situations where special accuracy near prescribed probability values is desired.

TABLE 3.3

Exact probabilities near .05 and .95, and errors of (3.10), (3.11), (3.12), (3.13) with α = .05

k	λ	$F_\lambda(k)$	(3.10)	(3.11)	(3.12)	(3.13)
4	1.85	.9600	-.0007	-.0008	-.0081	-.0070
4	1.97	.9500	+.0007	-.0048	-.0038	-.0060
4	2.08	.9400	+.0021	-.0091	+.0007	-.0049
4	9.51	.0400	-.0009	+.0030	-.0010	+.0018
4	9.15	.0500	+.0001	+.0013	+.0044	+.0032
4	8.86	.0600	+.0013	-.0006	+.0107	+.0048
15	9.73	.9600	-.0006	+.0007	-.0037	-.0021
15	10.04	.9500	+.0001	-.0010	-.0012	-.0015
15	10.30	.9400	+.0008	-.0029	+.0014	-.0009
80	97.42	.0400	-.0003	+.0006	-.0009	-.0001
80	96.35	.0500	+.0000	+.0001	+.0003	+.0002
80	95.44	.0600	+.0003	-.0005	+.0016	+.0006
120	102.45	.9600	-.0002	+.0004	-.0011	-.0004
120	103.49	.9500	+.0000	-.0001	-.0002	-.0002
120	104.39	.9400	+.0003	-.0007	+.0008	+.0001

4. BETTER APPROXIMATIONS

The preceding section dealt with simple normal deviates u with fixed continuity corrections. They have an error $\Phi(u) - F_\lambda(k)$ which is $O(\lambda^{-\frac{1}{2}})$. We recall that F_λ denotes the Poisson distribution function, that ξ is defined by $\Phi(\xi) = F_\lambda(k)$, and that all asymptotic relations hold for $\lambda \to \infty$ and bounded ξ.

The present section has five subsections, devoted to four types of normal approximations which have error $O(\lambda^{-1})$. The simple deviates have this property only for two special values of ξ for which the leading term of the error vanishes, but we shall now reduce the error to $O(\lambda^{-1})$ for all bounded ξ simultaneously. Sometimes this implies the existence of some special values of ξ for which the error is $O(\lambda^{-3/2})$. One can then try to choose the constants occurring in the deviate in such a way as to obtain special accuracy near preassigned probability levels, in the same way as this was done for simple deviates in section 3.

The *first type* with error $O(\lambda^{-1})$ uses transformations, other than square roots, of the Poisson variable. As these are rather awkward for hand calculations, we give no more than a brief outline of approximations with exponent 2/3 (including known results by ANSCOMBE, BLOM and MOORE) and an application of the WILSON-HILFERTY and WISHART approximations to the distribution of χ^2_ν.

The *second type* has structure $\Phi(u + R(v))$: to any of the simple deviates u discussed in section 3 one adds a correction term which is a function of a simple deviate v (one may take v = u in order to simplify the calculation).

The *third type* has structure $\Phi(u) + R(v)\ \phi(u)$: to the standard normal distribution function evaluated at a simple deviate u, one adds a correction as above, multiplied by the standard normal density function at u.

The *fourth type* is a simple deviate with "continuity corrections" a, b, c, d (see section 3) which are no longer constant, but functions of some simple deviate. For Poisson and gamma type this coincides with the second type, but for square root type it gives a new and rather promising result.

The *fifth type* is a linear combination of simple approximations.

4a. EXPONENT 2/3 OR 1/3 AND LOGARITHM

Our first type of deviate is based on the following expansion:

$$(4.1)\qquad (k+\gamma)^{2/3} = \lambda^{2/3}\,\{1 + \frac{2\xi}{3\lambda^{\frac{1}{2}}} + \frac{6\gamma-4}{9\lambda} + \frac{\xi^3 + (42-72\gamma)\xi}{324\lambda^{3/2}} + O(\lambda^{-2})\},$$

where γ is an arbitrary constant. Expansion (4.1) follows from (3.1), after explicit calculation of the $O(\lambda^{-3/2})$ term in the latter. If one puts $\zeta = 42-72\gamma$, one finds an error

$$(4.2)\qquad \phi(\xi)\ \lambda^{-1}\ (\xi^3 + \zeta\xi)/216 + O(\lambda^{-3/2})$$

for the following four approximations, in which $\beta = \gamma - 2/3$:

$$(4.3)\qquad \Phi(1.5(k+\gamma)^{2/3}\ (\lambda+4\beta)^{-1/6} - 1.5\lambda^{\frac{1}{2}}),$$

$$(4.4)\qquad \Phi(1.5(k+\gamma)^{2/3}\lambda^{-1/6} - 1.5\lambda^{\frac{1}{2}} - \beta\lambda^{-\frac{1}{2}}),$$

$$(4.5)\qquad \Phi(1.5(k+\gamma)^{2/3}\lambda^{-1/6} - 1.5(\lambda+4\beta/3)^{\frac{1}{2}}),$$

$$(4.6)\qquad \Phi(1.5\lambda^{-1/6}\{(k+\gamma)^{2/3} - (\lambda+\beta)^{2/3}\}).$$

ANSCOMBE (1953) advocates the use of $\underline{x}^{2/3}$ or $(\underline{x}+c)^{2/3}$ for normalization of a Poisson variable $\underline{x}$, because the transformed variable has skewness coefficient $O(\lambda^{-1})$. On the same occasion he proposes (4.6) with $\gamma = \frac{1}{2}$ as an approximation to $P[\underline{x} \leq k]$. BLOM (1954) suggests (4.4) with $\gamma = 1$, and MOORE (1957) uses the transformation $(\underline{x}+\frac{1}{4})^{2/3}$, which means $\gamma = \frac{3}{4}$ if a continuity correction of $\frac{1}{2}$ is added. Finally ANSCOMBE (1960), finding the three preceding proposals unsatisfactory, puts $\gamma = 2/3$; for this choice all four forms (4.3) - (4.6) are identical, as now $\beta = 0$.

We shall now compare the merits of the various values of γ. This is facilitated by Table 4.1, which gives $\beta = \gamma - 2/3$ and $\zeta = 42-72\gamma$. The leading term of the error, being proportional to $\xi^3 + \zeta\xi$, vanishes at $\xi = 0$ and for negative ζ also at $\pm(-\zeta)^{\frac{1}{2}}$. Thus the choice $\gamma = 2/3$ (which simplifies the formulae) provides asymptotic optimality at $\xi^2 = 6$. If one desires special accuracy at $\xi^2 = 4$ or $\xi^2 = 1$ one should take $\gamma = 23/36$ or $\gamma = 43/72$ respectively. It is easy to see that ANSCOMBE's second proposal $\gamma = 2/3$ is superior to his first idea $\gamma = \frac{1}{2}$: the asymptotic errors are proportional to $\xi^3 - 6\xi$ and $\xi^3 + 6\xi$ respectively, and the latter has a larger absolute value for any $\xi \neq 0$.

TABLE 4.1

Values of γ, $\beta = \gamma - 2/3$ and $\zeta = 42-72\gamma$, see text

γ	β	ζ	γ	β	ζ
1	1/3	-30	11/18	-1/18	-2
3/4	1/12	-12	43/72	-5/72	-1
2/3	0	-6	7/12	-1/12	0
23/36	-1/36	-4	1/2	-1/6	+6
5/8	-1/24	-3			

When a table of third roots of integers is available, it might be attractive to replace $(k+\gamma)^{2/3}$ by $\gamma(k+1)^{2/3} + (1-\gamma)k^{2/3}$; this gives the same error (4.2) with a different $0(\lambda^{-3/2})$ term. The same holds for $(k+\gamma_1)^{1/3}\ (k+\gamma_2)^{1/3}$ provided that $\gamma_1 + \gamma_2 = \gamma$: this seems to offer no special advantages.

The difference produced by such a replacement, and also the difference between the forms (4.3) - (4.6), is found entirely in the higher order terms of the error, the leading term being always as given in (4.2). We have not explicitly calculated the $0(\lambda^{-3/2})$ term for all cases. A numerical investigation demonstrated that (4.3) was nearly always less accurate than (4.4), (4.5) and (4.6), between which not much difference was found. Replacement of $(k+\gamma)^{2/3}$ by $\gamma(k+1)^{2/3} + (1-\gamma)k^{2/3}$ gave roughly the same accuracy for large values of λ, say $\lambda \geq 30$, but much worse results for smaller values. The number of exceptions to these rules was remarkably small, but of course one always finds some special pairs (k,λ) for which a usually bad approximation happens to be better than a generally accurate one.

The numerical study indicates that (4.4) with $\gamma = 5/8$, i.e.

$$(4.7) \qquad \Phi(1.5(k+.625)^{2/3}\,\lambda^{-1/6} - 1.5\lambda^{\frac{1}{2}} + \lambda^{-\frac{1}{2}}/24),$$

is generally accurate, even for smaller values of λ and k. Numerical information about its error is given in section 6. For small tails, say .02 or less, (4.4) with $\gamma = 23/36$ has an error roughly 4/5 of the error of (4.7), but it is less accurate for larger tails. A similar feature is shown by (4.4) with $\gamma = 2/3$ (β=0), which is relatively easy, accurate for small tails, but somewhat inaccurate for (say) tails of more than .02.

Because the calculation of (4.7) is rather unattractive, we shall proceed to seek simpler approximations with error $O(\lambda^{-1})$. As was announced in section 3b, one may combine the relation (1.4) $F_\lambda(k) = P[\chi^2_{2k+2} > 2\lambda]$ with the WILSON-HILFERTY (1931) result that the random variable $(\chi^2_\nu/\nu)^{1/3}$ is for $\nu \to \infty$ asymptotically normal with expectation $1 - 2/(9\nu)$ and variance $2/(9\nu)$. It follows that

$$(4.8) \qquad \Phi(3(k+1)^{\frac{1}{2}} - \tfrac{1}{3}(k+1)^{-\frac{1}{2}} - 3\{\lambda(k+1)^{\frac{1}{2}}\}^{1/3})$$

is an approximation to the Poisson distribution function $F_\lambda(k)$. In our opinion, the third root makes it somewhat unattractive for hand calculation, although it is less cumbersome than (4.7).

By straightforward series expansion of the deviate one finds that the error of (4.8) equals

$$(4.9) \qquad \phi(\xi)\,\lambda^{-1}\,(-\xi^3 + 3\xi)/108 + O(\lambda^{-3/2}).$$

This means asymptotic optimality at $\xi^2 = 3$, and it is indeed well known that the WILSON-HILFERTY approximation is very accurate near the .05 and .95 fractile, for which $|\xi| = 1.645$.

There is no simple modification of (4.8) which is in a similar way asymptotically optimal at other fractiles. One might try to replace e.g.

(k+1) by (k+h) or λ by $(\lambda+\theta)$ in (4.8), but that would mean h and/or θ proportional to $\lambda^{-\frac{1}{2}}$. This leads to a rather complicated expression.

The error (4.9) is asymptotically (-2) times the error (4.2) valid for the deviates involving $(k+\gamma)^{2/3}$, when one takes $\gamma = 5/8$ for asymptotic optimality at the same values $\xi = \pm\sqrt{3}$. Numerical experience confirms that (4.8) is usually less accurate than (4.7).

From a normal approximation to $P[\chi^2_\nu \leq \chi^2]$ published by WISHART (1956), one can derive that the Poisson distribution function $F_\lambda(k)$ is approximated by $\Phi(w - (k+1)^{-\frac{1}{2}}(w^2+2)/6)$, where $w = (k+1)^{\frac{1}{2}} \log\{(k+1)/\lambda\}$. This is a somewhat cumbersome expression. We omit the derivation of the expansion of its error, which is $\phi(\xi)\ \lambda^{-1}\ (-\xi^3 + \xi)/36 + O(\lambda^{-3/2})$. This is worse than for the deviates with exponent 1/3 or 2/3. WISHART gives some more correction terms, but their inclusion seems rather impractical for hand calculation. The error of $\Phi(w)$ is $\phi(\xi)\ \lambda^{-\frac{1}{2}}\ (\xi^2 + 2)/6 + O(\lambda^{-1})$; this is worse than the error of our simple square root type, which is easier to calculate.

4b. ADDITIVE CORRECTIONS TO DEVIATES

It follows from expansion (2.2) that

$$\text{(4.10)} \qquad \Phi(u_p + \lambda^{-\frac{1}{2}}\ (1 - u_p^2)/6), \text{ where } u_p = (k+\tfrac{1}{2}-\lambda)\lambda^{-\frac{1}{2}},$$

has error $O(\lambda^{-1})$. However, there is no special reason to use a corrected version of just u_p. If one starts from the more general simple deviate $u = (k+c-\lambda)\lambda^{-\frac{1}{2}}$, it is evident from (3.1) or (3.2) that a correction term $\lambda^{-\frac{1}{2}}\ (4-6c - \xi^2)/6$ should be added. Note that the value of c is irrelevant, as the terms $c\lambda^{-\frac{1}{2}}$ cancel. As ξ is unknown, we replace it by some simple deviate v (we may take v = u in order to simplify calculations), and the error remains $O(\lambda^{-1})$, as $v-\xi = O(\lambda^{-\frac{1}{2}})$. Continuing along these lines one finds the possible choices

$$\text{(4.11)} \qquad \Phi(u + (\lambda+\theta)^{-\frac{1}{2}}\ Q_u(v)),$$

$$\text{(4.12)} \qquad \Phi(u + (k+h)^{-\frac{1}{2}}\ Q_u(v)),$$

where u and v can be any simple deviate, θ and h are constants, and Q_u is a polynomial:

$$Q_u(v) \equiv (4-6c-v^2)/6 \text{ when } u = (k+c-\lambda)\lambda^{-\frac{1}{2}} \text{ (Poisson type)},$$

$$(4.13) \qquad Q_u(v) \equiv (v^2+8-12a+12b)/12 \text{ when } u = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}} \text{ (square root type)},$$

$$Q_u(v) \equiv (v^2-1-3d)/3 \text{ when } u = (k+1+d-\lambda)(k+1)^{-\frac{1}{2}} \text{(gamma type)}.$$

From a tedious calculation, not reproduced here, it follows that any approximation (4.11) or (4.12) has an error of the form

$$(4.14) \qquad \phi(\xi)\ \lambda^{-1}\ (e_3\xi^3 + e_1\xi) + O(\lambda^{-3/2}),$$

where the coefficient e_3 depends only on the types of the deviates u and v and on the choice between $(\lambda+\theta)^{-\frac{1}{2}}$ and $(k+h)^{-\frac{1}{2}}$; the coefficient e_1 depends also on the constants occurring in u and v. The values of θ or h have no influence on the leading term of the error, but do affect the $O(\lambda^{-3/2})$ term. The numerical values of the coefficients e_3 and e_1 are found in Table 4.2.

Now only four of the eighteen possibilities listed in Table 4.2 deserve further consideration; the other fourteen can be discarded. We shall illustrate the reason for this by comparing (4.11) and (4.12) for both u and v of Poisson type. To any choice $c = c_0$ for (4.11) there corresponds the choice $c = (2c_0+3)/5$ for (4.12), which yields $|-5\xi^3/72 - c_0\xi/3 + 7\xi/36| = 5|\xi^3/72 + (2c_0/5+3/5)\ \xi/6 - 5\xi/36|$: the leading term of the error for (4.11) is uniformly in ξ five times larger than the leading term of the error for a corresponding version of (4.12), see Table 4.2.

As a similar manipulation of the coefficient e_1 of ξ is always possible, we retain only the four cases in Table 4.2 for which the coefficient e_3 of ξ^3 is minimal: for (4.12) v Poisson and u Poisson or square root, and for (4.11) v square root and u Poisson or square root. These four have the same asymptotic error for suitable a, b and c.

As the use of different types for u and v tends to complicate the calculations, we discard (4.12) with v Poisson and u square root, and consider for (4.11) with v square root and u Poisson only $\theta = 0$, which permits simplification. There remain:

(4.15) $\Phi(u + (k+h)^{-\frac{1}{2}}(4-6c-u^2)/6)$ with $u = (k+c-\lambda)\lambda^{-\frac{1}{2}}$;

(4.16) $\Phi(u + (\lambda+\theta)^{-\frac{1}{2}}(u^2+8-12a+12b)/12)$ with $u = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}$;

(4.17) $\Phi(k - \lambda + (4-v^2)/6)\lambda^{-\frac{1}{2}})$ with $v = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}$.

TABLE 4.2

Values of the polynomial $(e_3\xi^3 + e_1\xi)$, occurring in the leading term of the error of (4.11) and (4.12), for different types of the simple deviate u and v. As before, we put $\delta = a-b$

		v Poisson type	v square root type	v gamma type
u Poisson type	(4.11)	$-\frac{5\xi^3}{72} - \frac{c\xi}{3} + \frac{7\xi}{36}$	$\frac{\xi^3}{72} - \frac{\delta\xi}{3} + \frac{7\xi}{36}$	$\frac{7\xi^3}{24} - \frac{d\xi}{3} - \frac{5\xi}{36}$
	(4.12)	$\frac{\xi^3}{72} + \frac{c\xi}{6} - \frac{5\xi}{36}$	$\frac{7\xi^3}{72} - \frac{\delta\xi}{3} + \frac{c\xi}{2} - \frac{5\xi}{36}$	$\frac{13\xi^3}{72} - \frac{d\xi}{3} + \frac{c\xi}{2} - \frac{5\xi}{36}$
u square root type	(4.11)	$\frac{\xi^3}{18} - \frac{a\xi}{2} + \frac{c\xi}{6} + \frac{7\xi}{36}$	$\frac{\xi^3}{72} - \frac{\delta\xi}{3} - \frac{b\xi}{2} + \frac{7\xi}{36}$	$-\frac{\xi^3}{36} + \frac{d\xi}{6} - \frac{a\xi}{2} + \frac{13\xi}{36}$
	(4.12)	$\frac{\xi^3}{72} - \frac{b\xi}{2} + \frac{c\xi}{6} - \frac{5\xi}{36}$	$-\frac{\xi^3}{36} + \frac{\delta\xi}{6} - \frac{b\xi}{2} - \frac{5\xi}{36}$	$-\frac{5\xi^3}{72} + \frac{d\xi}{6} - \frac{b\xi}{2} + \frac{\xi}{36}$
u gamma type	(4.11)	$\frac{11\xi^3}{36} - \frac{d\xi}{2} + \frac{2c\xi}{3} - \frac{29\xi}{36}$	$\frac{5\xi^3}{36} - \frac{d\xi}{2} + \frac{2\delta\xi}{3} - \frac{29\xi}{36}$	$-\frac{\xi^3}{36} + \frac{d\xi}{6} - \frac{5\xi}{36}$
	(4.12)	$\frac{5\xi^3}{36} + \frac{2c\xi}{3} - \frac{23\xi}{36}$	$-\frac{\xi^3}{36} + \frac{2d\xi}{3} - \frac{23\xi}{36}$	$-\frac{7\xi^3}{36} + \frac{2d\xi}{3} + \frac{\xi}{36}$

The freedom to choose c in (4.15) or a and b in (4.16) and (4.17) can be exploited to make the leading term of the error vanish for preassigned values of ξ. The constants h or θ affect only terms of higher order. As an illustration we consider (4.15), which has an error

$$\begin{aligned} &\phi(\xi)\ [\lambda^{-1}\ (\xi^3 + 12c\xi-10\xi)/72 + \\ (4.18)\qquad &+ \lambda^{-3/2}\ (-111\xi^4 + 536\xi^2-1170c\xi^2+540h\xi^2 + \\ &+ 832-720c-2160h-1080c^2+3240ch)/6480] + O(\lambda^{-2}). \end{aligned}$$

It follows e.g. that $c = \frac{1}{2}$ gives asymptotic optimality at $\xi^3 - 4\xi = 0$, i.e. $\xi = \pm 2$, tails of .023, and $\xi = 0$, the median. The error is then $\phi(\xi)\ \lambda^{-1}\ (\xi^3 - 4\xi)/72 + O(\lambda^{-3/2})$; the choice of h only affects the $O(\lambda^{-3/2})$ term and cannot be used to improve the asymptotic accuracy for any ξ which is not - 2, 0 or 2. Therefore we choose h in such a way that for $|\xi| = 2$ the term of order $\lambda^{-3/2}$ vanishes too: we attach more importance to optimality for tails of .023 than to optimality near the median. Substitution of $c = \frac{1}{2}$ and $\xi^2 = 4$ into (4.18) shows that one must take $h = 59/54$. Numerical investigation shows that $c = \frac{1}{2}$, $h = 1$ is not much worse, and it simplifies the calculation of (4.15).

Similar considerations show that $c = 2/3$, $h = 53/54$ makes (4.15) asymptotically optimal at $\xi^2 = 2$. Numerical investigation shows that this approximation is generally accurate for probabilities between .05 and .95, and that $h = 1$ is hardly less accurate (and easier for calculation).

The errors of (4.16) and (4.17) can be written in a form similar to (4.18). We shall not give the explicit formulae, but just give some comment on the best choices of the constants and the numerical results.

For (4.16) one finds that $a = .59$, $b = .31$ and $\theta = 0$ give good accuracy near $\xi^2 = 4$, and the same a and θ with $b = .16$ lead to close approximation of the middle part. The first mentioned triple (a,b,θ) gives in (4.16) a result which is for small λ (say $.5 \leq \lambda \leq 20$) somewhat more accurate than the versions of (4.15) or (4.17) with the same asymptotic error. BLOM (1954) mentions (4.16) with $a = 1$, $b = 0$ and $\theta = 0$.

In (4.17) the simple values $a = \frac{3}{4}$, $b = 0$ or $a = 2/3$, $b = 0$ are almost equal to the complicated fractional values giving asymptotic optimality for $\xi^2 = 4$ or $\xi^2 = 2$ respectively. They are roughly as accurate as the corresponding versions of (4.15).

4c. ADDITIVE CORRECTIONS TO PROBABILITIES

The third class of $O(\lambda^{-1})$ approximations has the form

(4.19) $$\Phi(u) + (\lambda+\theta)^{-\frac{1}{2}} Q_u(v)\ \phi(u),$$

or

(4.20) $$\Phi(u) + (k+h)^{-\frac{1}{2}} Q_u(v)\ \phi(u),$$

where u and v are simple deviates, Q_u is defined in (4.13), θ and h are constants and ϕ denotes the standard normal density function. One special case, namely (4.19) with $u = v = (k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$ and $\theta = 0$, occurs in the literature: for its absolute error, CHENG (1949) gives an upper bound depending on λ and valid for all k simultaneously. When the same approximation is used to find $F_\lambda(k_2) - F_\lambda(k_1)$, a similar bound is found in MAKABE & MORIMURA (1955). As is shown in Table 4.3, such bounds are not very informative about actual errors.

TABLE 4.3

Numerical values of the error bounds of CHENG and MAKABE & MORIMURA, compared with the largest error actually observed

λ	CHENG's error bound	actual largest error	M & M's error bound	actual largest error
1	.2490	.0153	.5490	.0171
4	.0325	.0069	.0552	.0107
10	.0103	.0024	.0106	.0041
60	.0014	.0003	.0010	.0006

We now return to the general formulae (4.19) and (4.20). Their error has the form

$$(4.21) \qquad \phi(\xi)\ \lambda^{-1}\ (e_5\xi^5 + e_3\xi^3 + e_1\xi) + O(\lambda^{-3/2});$$

the type of u determines the coefficient e_5, whereas e_3 and e_1 depend also on the choice of the type of v and of the constants occurring in u and v. The constants θ or h only have influence on the $O(\lambda^{-3/2})$ term.

The value of e_5 is 1/288, 1/72 and 1/18 when u has square root type, Poisson type and gamma type respectively. In the same way as for the second class of approximations, it follows that one can restrict attention to square root type u: the error of any formula with Poisson or gamma type u is asymptotically 4 or 16 times larger than a formula with square root type u and suitable choice of the constants a and b. Furthermore, only the case v = u is considered in order to avoid the computation of two simple deviates.

This means consideration of

$$(4.22) \qquad \Phi(u) + (\lambda+\theta)^{-\frac{1}{2}}(u^2+8-12a+12b)\ \phi(u),$$

or

$$(4.23) \qquad \Phi(u) + (k+h)^{-\frac{1}{2}}(u^2+8-12a+12b)\ \phi(u),$$

where $u = 2\ (k+a)^{\frac{1}{2}} - 2\ (\lambda+b)^{\frac{1}{2}}$. Now (4.22) has error (put δ = a-b)

$$(4.24) \qquad \phi(\xi)\lambda^{-1}(\xi^5+20\xi^3-24\delta\xi^3+120\xi-144b\xi-288\delta\xi+144\delta^2\xi)/288 + O(\lambda^{-3/2}),$$

and (4.23) has error

$$(4.25) \qquad \phi(\xi)\lambda^{-1}(\xi^5+8\xi^3-24\delta\xi^3+24\xi-144b\xi-144\delta\xi+144\delta^2\xi)/288 + O(\lambda^{-3/2}).$$

The polynomial of fifth degree in (4.24) and (4.25) can be analysed, numerically or by consideration of its zeroes and relative extremes. One finds that a = .9, b = - .2 in (4.24), and a = $\frac{1}{2}$, b = -1/8 in (4.25), make the polynomials reasonably small for all $|\xi| < 2.1$. This is illustrated in Fig. 4.1, which gives also some corresponding polynomials: $(\xi^3 - 3\xi)/216$ for a 2/3-power transformation, $(-\xi^3 + 3\xi)/108$ for the WILSON-HILFERTY (4.8) approximation and $(\xi^3 - 4\xi)/72$ for (4.15), (4.16) or (4.17) with appropriate choices of a, b or c.

The values of θ in (4.22) and h in (4.23) have no influence on the leading term of the error. Numerical investigation shows that one may take $\theta = h = 0$.

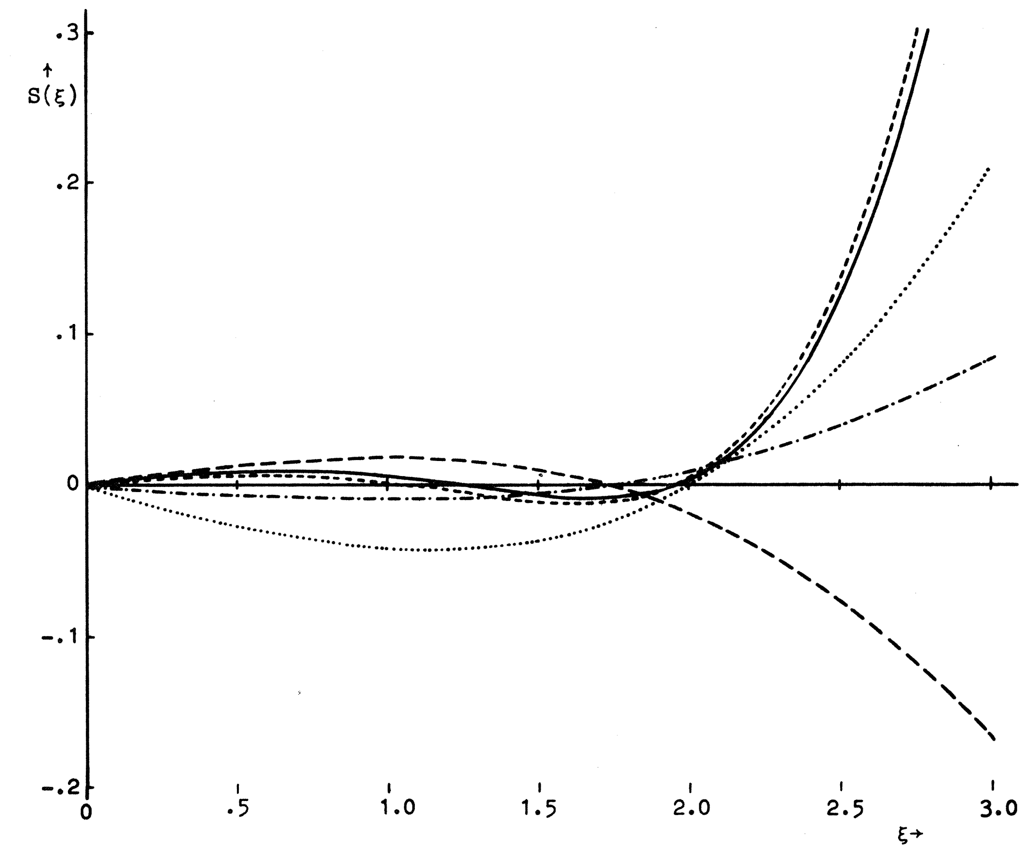

Fig. 4.1 Polynomials $S(\xi)$ such that the error of certain normal approximations to the Poisson distribution equals $\phi(\xi)\lambda^{-1}S(\xi) + O(\lambda^{-3/2})$:
——— (4.25) for $a = \frac{1}{2}$, $b = -1/8$; ---------- (4.24) for $a = .9$, $b = -.2$;
-·-·-·- (4.2) 2/3 power, $\gamma = 5/8$, $\zeta = -3$; – – – – (4.9) Wilson-Hilferty;
.... $S(\xi) = (\xi^3-4\xi)/72$ e.g. for (4.15) with $c = \frac{1}{2}$.

The Figure suggests that (4.24) and (4.25), with the mentioned choices of a and b, are usually smaller than the errors of the other types, unless $|\xi|$ is large. Numerical investigation shows that the difference in accuracy is somewhat less than expected, and the behaviour for large $|\xi|$ is worse than expected. This is presumably due to the terms of higher order in the error, which now include a term $\xi^8\lambda^{-3/2}$, whereas the other types had at most $\xi^4\lambda^{-3/2}$. We may conclude that (4.22) and (4.23) do not provide enough accuracy for the substantial amount of calculation required by them.

4d. VARIABLE CONTINUITY CORRECTIONS

We now turn to the fourth method of obtaining approximations with error $O(\lambda^{-1})$. The constants a, b, c, d occurring in the simple deviates of section 3 can be replaced by functions A(v), B(v), C(v), D(v) of some simple deviate v, and it is immediately clear from the series expansions that one should consider

(4.26) $\quad \Phi((k+C(v)-\lambda)\lambda^{-\frac{1}{2}}) \qquad$ with $C(v) \equiv (4-v^2)/6$,

(4.27) $\quad \Phi((k+1+D(v)-\lambda)(k+1)^{-\frac{1}{2}}) \qquad$ with $D(v) \equiv (v^2-1)/3$,

(4.28) $\quad \Phi(2\{k+A(v)\}^{\frac{1}{2}} - 2\{\lambda+B(v)\}^{\frac{1}{2}}) \qquad$ with $A(v) - B(v) \equiv (8+v^2)/12$.

But now (4.26) coincides with (4.11) if one puts there $u = (k+c-\lambda)\lambda^{-\frac{1}{2}}$ and $\theta = 0$; similarly (4.17) equals (4.12) with $u = (k+1+d-\lambda)(k+1)^{-\frac{1}{2}}$ and $h = 1$. In their present form they may be slightly easier for computation.

Formula (4.28) is not of the form (4.11) or (4.12), as A and B occur under a radical, unlike C and D which are additive corrections. This means that a new approximation with error $O(\lambda^{-1})$ is obtained for any A and B with $A(v) - B(v) \equiv (8+v^2)/12$. By series expansion one finds, expressing A in terms of B, that the error of (4.28) equals:

(4.29) $\quad \phi(\xi)\,\lambda^{-1}\,\{\xi^3 + (12c-10-36B)\xi\}/72 + O(\lambda^{-3/2})$

when $v = (k+c-\lambda)\lambda^{-\frac{1}{2}}$,

(4.30) $\phi(\xi)\lambda^{-1}\{-5\xi^3 + (12d+2-36B)\xi\}/72 + O(\lambda^{-3/2})$
when $v = (k+1+d-\lambda)(k+1)^{-\frac{1}{2}}$,

(4.31) $\phi(\xi)\lambda^{-1}\{-2\xi^3 + (12a-12b-10-36B)\xi\}/72 + O(\lambda^{-3/2})$
when $v = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}$.

As B is allowed to be any polynomial in v, it follows that a suitable choice of B can give approximation (4.28) an error of order $\lambda^{-3/2}$. This point is further investigated in section 5, which deals with that kind of approximations. Here we are content with order λ^{-1}, which means that we can choose B in such a way that the resulting expression (4.28) is relatively simple. The choices

$$B(v) \equiv 0,\ A(v) \equiv (8+v^2)/12,$$
(4.32) $$B(v) \equiv -(8+v^2)/12,\ A(v) \equiv 0,$$
$$B(v) \equiv -(-4+v^2)/12,\ A(v) \equiv 1$$

will be considered. For each of them we are still free to choose the deviate v and the constants a, b, c, d that may occur in v. However, by the type of argument used for the second and third class of approximations, it follows from (4.29), (4.30), (4.31) that for $B(v) \equiv 0$ one should take $v = (k+c-\lambda)\lambda^{-\frac{1}{2}}$ and for $A(v) \equiv 0$ or $A(v) \equiv 1$ one should take $v = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}$, because that gives an error

(4.33) $$\phi(\xi)\,\lambda^{-1}\,(\xi^3/72 + e_1\xi) + O(\lambda^{-3/2}),$$

with e_1 dependent on c or on a and b, whereas all other combinations have a larger coefficient of ξ^3, which means that their asymptotic error is uniformly in ξ larger than necessary.

From theoretical and numerical study of the errors it follows that for

(4.34) $$\Phi(2\{k + (v^2+8)/12\}^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}),\ v = (k+c-\lambda)\lambda^{-\frac{1}{2}},$$

one must take $c = \frac{1}{2}$ or $c = 2/3$ for asymptotic optimality at $\xi^2 = 4$ or

$\xi^2 = 2$ respectively. For

$$(4.35) \qquad \Phi(2\{k+1\}^{\frac{1}{2}} - 2\{\lambda + (4-v^2)/12\}^{\frac{1}{2}}),\ v = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}},$$

the choices are $a = 3503/2160$, $b = 263/2160$ ($\xi^2 = 4$) and $a = 2047/1080$, $b = 247/1080$ ($\xi^2 = 2$) respectively. Replacement by $a = 13/8$, $b = 1/8$ and $a = 2$, $b = 1/3$ simplifies the calculations without serious loss of accuracy. The deviates (4.28) with $A(v) \equiv 0$ are nearly always less accurate.

4e. LINEAR COMBINATIONS

The expansions of the simple approximations discussed in section 3 suggest that an error $O(\lambda^{-1})$ is also achieved by suitable linear combinations like

$$(4.36) \qquad \Phi(\{(k+c-\lambda)\lambda^{-\frac{1}{2}} + 2u_{a,b}\}/3),$$

where $u_{a,b}$ is given by (3.7) with $a-b = 1-\frac{1}{2}c$. Other possibilities like $\{\Phi((k+c-\lambda)\lambda^{-\frac{1}{2}}) + 2\Phi(u_{a,b})\}/3$, or combinations of Poisson and gamma type, are usually less accurate.

COMPARISON OF BETTER APPROXIMATIONS

Within the class with error $O(\lambda^{-1})$, the 2/3 power approximations like (4.7) are nearly always best, followed by WILSON-HILFERTY (4.8). Among the others accuracy varies only slightly, although the third type $\Phi(u) + R(v)\phi(u)$ is definitely inaccurate for small tails. One could use (4.36) with $c = \frac{1}{2}$ or (4.34) with $c = \frac{1}{2}$, which are relatively simple. If third roots could be easily obtained and accuracy was important, one might use (4.7) instead. However, the question is somewhat academic, as the next section presents a deviate which is relatively easy to calculate and generally more accurate than any $O(\lambda^{-1})$ approximation. Loosely speaking we might say that the approximations of the present section provide not enough accuracy for the amount of calculation involved, when they are compared to the new formula (5.10).

5. VERY ACCURATE APPROXIMATIONS

A normal approximation with error $O(\lambda^{-3/2})$ can only be obtained at cost of somewhat lengthy calculations. This section presents a brief discussion of some very accurate but cumbersome formulae.

PEIZER & PRATT (1968) propose two normal approximations to the whole class of beta, gamma and related distributions. For the Poisson case they use

$$(5.1) \qquad \Phi(\{k - \lambda + \tfrac{2}{3} + \tfrac{\varepsilon}{k+1}\} \{1 + T(\tfrac{k+\frac{1}{2}}{\lambda})\}^{\frac{1}{2}} \lambda^{-\frac{1}{2}}),$$

where $T(z) = (1 - z^2 + 2z \log z)\,(1-z)^{-2}$ and $T(1) = 0$ by continuity; they take $\varepsilon = 0$ for simplicity or $\varepsilon = .02$ for more accurate results. Their paper contains a table of T as a function of z.

From a straightforward but somewhat lengthy calculation of the expansion of the deviate, one obtains that the error of (5.1) is

$$(5.2) \qquad \phi(\xi)\, \lambda^{-3/2} (-\xi^2 + 1620\varepsilon-32)/1620 + O(\lambda^{-2}).$$

Thus for $\varepsilon = 0$ the error is asymptotically minimal at $\xi = 0$ (for which value the leading term does not vanish). For $\varepsilon = .02$ the error is locally $O(\lambda^{-2})$ for $\xi^2 = .4$ (tails of .26), and for $\varepsilon = .022$ the optimum lies at $\xi^2 = 3.64$ (tails of .028). We therefore propose to use (5.1) with $\varepsilon = .022$ whenever general accuracy near the customary significance levels is desired. Choices of ε especially designed for accuracy at special probability values α and $1-\alpha$ are listed in Table 5.1.

TABLE 5.1

Choices of ε which make (5.1) asymptotically optimal near the α and $1-\alpha$ fractiles.

α	.1	.05	.025	.01	.005
ε	.02077	.02142	.02212	.02309	.02385

The approximation (5.1) is extremely accurate; unlike most other approximations it remains so for very small tails. However, for $k = 0$ it is considerably worse than for larger values of k. This is not too serious, as explicit evaluation of $F_\lambda(0) = e^{-\lambda}$ is not difficult. Numerical values of errors are given in section 6. We refer to PEIZER & PRATT (1968) for further details.

The simple square root deviate $u_{a,b}$ (3.7) can also be modified into an $O(\lambda^{-3/2})$ approximation, if one replaces a and b by functions $A(v)$ and $B(v)$ of some simple deviate v. As already stated in section 4d, this gives the deviate (4.28); its error for the three possible types of v is (4.29), (4.30) or (4.31). The leading term of this error vanishes for a suitable choice of $B(v)$, followed by solution of $A(v)$ from the condition $A(v)-B(v) \equiv$ $\equiv (8+v^2)/12$.

From calculations not reproduced here one finds

(5.3) $\Phi(2\{k + (2v^2+7+6c)/18\}^{\frac{1}{2}} - 2\{\lambda + (v^2-10+12c)/36\}^{\frac{1}{2}})$,

with $v = (k+c-\lambda)\lambda^{-\frac{1}{2}}$ and error

$\phi(\xi)\ \lambda^{-3/2}\ (-3\xi^4 + 13\xi^2 + 56-360c+270c^2)/3240 + O(\lambda^{-2})$;

(5.4) $\Phi(2\{k + (v^2+14+12\delta)/36\}^{\frac{1}{2}} - 2\{\lambda + (-v^2-5+6\delta)/18\}^{\frac{1}{2}})$,

with $v = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}$, $\delta = a-b$, and error

$\phi(\xi)\ \lambda^{-3/2}\ (111\xi^4 + 824\xi^2-1080\delta\xi^2-2160b\xi^2 +$

$+ 448-2880\delta+2160\delta^2)/25920 + O(\lambda^{-2})$;

(5.5) $\Phi(2\{k + (-v^2+13+6d)/18\}^{\frac{1}{2}} - 2\{\lambda + (-5v^2+2+12d)/36\}^{\frac{1}{2}})$,

with $v = (k+1+d-\lambda)(k+1)^{-\frac{1}{2}}$ and error

$\phi(\xi)\ \lambda^{-3/2}\ (-3\xi^4 - 77\xi^2 - 34+180d-270d^2)/3240 + O(\lambda^{-2})$.

From the asymptotic errors one obtains that (5.3), with a low coefficient of ξ^4 and a contribution of ξ^2 partially compensating it, can be expected to be better that the other two. The leading term of the error of

(5.3) vanishes at $\xi^2 = 4$ for $c = (2 \pm \sqrt{2})/3$, i.e. 1.14 and .20, and at $\xi^2 = 2$ for $c = (6 \pm \sqrt{15})/9$, i.e. 1.10 and .24. However, the influence of the $O(\lambda^{-2})$ term will be considerable unless λ is very large.

A third possibility is a simple deviate with two corrections, either "within Φ" (second type of section 4) or proportional to ϕ (third type of section 4). This leads anyhow to complicated calculations. Calculation of the asymptotic errors shows that the simple deviate should be of square root type, but even then the coefficient of ξ^4 has a larger absolute value than 1/1080, which was reached for (5.3). We have selected for numerical study

$$(5.6) \qquad \Phi(u + (\lambda+\theta)^{-\frac{1}{2}} (u^2+8-12\delta)/12 + (\lambda+\theta)^{-1} (-u^3-14u+24\delta u+36bu)/72),$$

$$(5.7) \qquad \Phi(u + (k+h)^{-\frac{1}{2}} (u^2+8-12\delta)/12 + (k+h)^{-1} (u^3+5u-6\delta u+18bu)/36),$$

$$(5.8) \qquad \begin{aligned} &\Phi(u) + \phi(u) \{(\lambda+\theta)^{-\frac{1}{2}} (u^2+8-12\delta)/12 + \\ &+ (\lambda+\theta)^{-1} (-u^5-20u^3+24\delta u^3-120u+144bu+288\delta u-144\delta^2 u)/288\}, \end{aligned}$$

$$(5.9) \qquad \begin{aligned} &\Phi(u) + \phi(u) \{(k+h)^{-\frac{1}{2}} (u^2+8-12\delta)/12 + \\ &+ (k+h)^{-1} (-u^5-8u^3+24\delta u^3-24u+144bu+144\delta u-144\delta^2 u)/288\}, \end{aligned}$$

where $u = 2(k+a)^{\frac{1}{2}} - 2(\lambda+b)^{\frac{1}{2}}$ and $\delta = a-b$; we are free to choose the constants a, b, θ and h.

Numerical comparison of these approximations leads to the following conclusions. The PEIZER-PRATT formula (5.1) is extremely accurate, with an exception for left hand tails (mainly k=0) when roughly $\lambda \leq 5$. The choice $\varepsilon = .022$ is better than $\varepsilon = .02$, which in turn is better than $\varepsilon = 0$. Especially for left tails λ must be rather large (at least something like 20) in order to obtain that the probability value of optimal accuracy comes very close to the value α corresponding to ε given in Table 5.1; for such values of λ the relative tail error, defined in (2.3) of Chapter I, is anyhow less than .1 per cent. for all tails of at least .001. For roughly

$\lambda \geq 100$ the choice of ε has very little influence.

Of the corrected square root deviates, (5.3) is decidedly superior to (5.4) and (5.5). The choice $c = 1/6$ gives both general accuracy and a simple expression: it leads to

$$(5.10) \qquad \Phi(2\{k + (t+4)/9\}^{\frac{1}{2}} - 2\{\lambda + (t-8)/36\}^{\frac{1}{2}}),$$

with $t = (k - \lambda + 1/6)^2\lambda^{-1}$ and error

$$\phi(\xi)\ \lambda^{-3/2}\ (-6\xi^4 + 26\xi^2 + 7)/6480 + O(\lambda^{-2}).$$

This deviate is usually slightly less accurate than the PEIZER-PRATT deviate (5.1) with $\varepsilon = .022$; this holds for the leading terms of the asymptotic errors, and also for most numerical examples. However, (5.10) is still extremely accurate (see the numerical information in section 6), and in our opinion it is much easier to compute.

Of the deviates with two corrections, a good choice is (5.6) with $a = \frac{3}{4}$ and $b = \theta = 0$, i.e.

$$(5.11) \qquad \Phi(u + \lambda^{-\frac{1}{2}}(u^2-1) + \lambda^{-1}(-u^3+4u)/72), \text{ with } u = 2(k+\tfrac{3}{4})^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}};$$

for small λ it is sometimes better and sometimes worse than (5.10), but for (say) $\lambda \geq 5$ it is decidedly less accurate near the customary significance levels and also in most of the middle part. For computation it is somewhat less handy than (5.10).

Summarizing we may say that (5.10) offers much accuracy for a relatively low amount of computation. The more cumbersome (5.1) with $\varepsilon = .022$ constitutes an improvement for right hand tails for roughly $\lambda \leq 30$, and for the middle part for roughly $\lambda \leq 5$. Both approximations make a relative error of less than 1 per cent. in any tail $\geq .001$ whenever $\lambda \geq 7$, and are still rather useful when λ is as small as 1.

An error of order $\lambda^{-3/2}$ is also achieved by a special linear combination of all three simple deviates, viz.

$$(5.12) \qquad \Phi(\{2(k+\tfrac{3}{4}-\lambda)\lambda^{-\frac{1}{2}} - (k+\tfrac{1}{2}-\lambda)(k+1)^{-\frac{1}{2}} + 16(k+\tfrac{5}{8})^{\frac{1}{2}} - 16\lambda^{\frac{1}{2}}\}/9).$$

However, (5.10) and (5.1) are more accurate and less cumbersome.

6. GENERAL ADVICE AND NUMERICAL INFORMATION

This section opens with a general advice on normal approximations to the Poisson distribution function (Table 6.1). We recall that a simple recommendation can only be given after an evaluation of accuracy and labouriousness involving some subjective elements, and disregarding some exceptions. The principles guiding our choice have been explained in the summary, and more fully in Chapter I. In our opinion, the approximations selected for Table 6.1 combine a maximum of accuracy with a minimum of effort.

The remainder of this section contains the results of an effort to condense a huge amount of numerical information on errors into a few pages. In Tables 6.2 and 6.3 we give the relative tail error defined in I (2.3), for some approximations selected from sections 3, 4 and 5. For the Poisson parameter λ we have chosen the values .5, 2, 10, 30 and 200. The Tables show the general superiority of the square root approximation (3.7) compared to the classical approximation (3.1). One can also verify to what extent somewhat laborious approximations like (5.1) or (4.4) are better than the improved square root approximation (5.10). This formula (5.10) is the version of (5.3) with $c = 1/6$. It is recommended in Table 6.1, as it is rather accurate and relatively simple.

The section ends with three graphs in which the Poisson parameter λ is given on the vertical (logarithmic) scale, the Poisson distribution function $F_\lambda(k) = P$ on the horizontal (normal probability) scale, and points of equal (absolute) error P^*-P have been joined. Each graph pertains to one of the three square root approximations P^* that we have recommended. For a clear presentation continuous curves of constant error are sketched in the Figures, but actually the points in the (P,λ) plane for which the error has a fixed value form a discrete set. Indeed there is a countable number of possible values of $P = F_\lambda(k)$ for fixed λ, because k can only assume the values 0, 1, 2,... . As an illustration, the possible values corresponding to k = 0, 1 and 2 are sketched in Fig. 6.1. The complete set of possible values coincides with the Poisson nomogram (pages 28 and 29). There are no curves of constant error in the lower left hand corner of the Figures, because the points marked o, corresponding to k = 0, indicate the lowest possible values

(text continued on page 69)

TABLE 6.1. Advice for normal approximation to the Poisson distribution function $F_\lambda(k) = \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j!$. In most statistical applications, accurate approximation to probabilities between .005 and .05 or between .95 and .995 will be essential. In such cases, one may use the suggestion marked "for tails". Φ denotes the standard normal distribution function I (2.1).

For quick work, use (3.7)

$\Phi(2\{k+1\}^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$ for tails

$\Phi(2\{k+\frac{3}{4}\}^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$ for values between .06 and .94 (or .05 and .93 for roughly $\lambda < 15$).

Never use $(k-\lambda)\lambda^{-\frac{1}{2}}$ or $(k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}$, because the square root approximations displayed above are essentially more accurate and just as simple.

For special accuracy near probabilities α *and* $1-\alpha$, use (3.10)

$$\Phi(2\{k + (\xi_\alpha^2+11)/18\}^{\frac{1}{2}} - 2\{\lambda - (\xi_\alpha^2+2)/36\}^{\frac{1}{2}}),$$

where ξ_α is defined by $\Phi(\xi_\alpha) = 1-\alpha$; for numerical values of the polynomials in ξ_α cf. Table 3.1 on page 42.

For accurate approximation, use (5.10)

$$\Phi(2\{k + (t+4)/9\}^{\frac{1}{2}} - 2\{\lambda + (t-8)/36\}^{\frac{1}{2}}),$$

where $t = (k - \lambda + 1/6)^2/\lambda$.

TABLE 6.2. Event $x \leq k$ or $x \geq k+1$, exact Poisson probability and relative tail error in per cent. for some normal approximations
Example : for $F_2(1)$ = .4060, approximation (5.3) with c=0 has a relative error of +.60 per cent., it gives 1.0060×.4060 = .4084

Event	Proba-bility	Peizer-Pratt=(5.1) ε=0	ε=.02	ε=.022	square root=(5.3) c=0	c=1/6	c=1/2	(5.11)	(4.4) γ=2/3	(4.4) γ=5/8	(4.34) c=1/2	(3.7) a=1,b=0	(3.1) c=0	(3.7) a=3/4,b=0	(3.1) c=1/2
λ = .5															
$x \geq 1$	.3935	+3.40	+.62	+.34	-8.80	-3.29	+2.63	+2.35	+4.53	+4.05	+5.07	-29.09	+93.22	-4.62	+27.07
$x \geq 2$	.0902	+2.08	+.01	-.20	-15.65	-7.89	+1.12	-6.66	+6.15	+2.56	+8.50	-12.81	+165.79	+20.90	-12.81
$x \geq 3$	.0144	+1.83	+.11	-.06	-8.39	+.43	+10.11	-10.80	+2.43	-3.90	-12.19	+40.31	+17.79	+98.50	-83.74
$x \geq 4$	.0018	+1.83	+.34	+.19	+20.01	+27.71	+33.52	+1.95	-7.18	-15.39	-50.42	+177.33	-88.38	+297.95	-99.36
λ = 2															
$x \leq 0$	.1353	-4.38	-1.71	-1.44	-.27	-1.76	-3.17	-.68	+.00	-2.15	+2.78	+50.53	-41.89	+.83	+6.71
$x \leq 1$	.4060	-.85	-.14	-.07	+.60	-.04	-.75	-.65	-.47	-.66	-.05	+23.15	-40.95	+5.30	-10.88
$x \geq 3$	.3233	+.54	+.04	-.01	-.90	-.28	+.42	+.26	+1.11	+.81	+1.62	-18.81	+54.64	-3.28	+11.91
$x \geq 4$	.1429	+.52	+.02	-.03	-1.43	-.64	+.30	-.56	+1.78	+.93	+2.80	-15.53	+67.80	+3.67	+1.08
$x \geq 5$	.0527	+.52	+.03	-.02	-1.70	-.72	+.49	-1.53	+1.88	+.39	+1.78	-4.81	+49.37	+19.56	-26.78
$x \geq 6$	.0166	+.53	+.05	+.00	-1.14	+.02	+1.49	-1.92	+1.08	-1.07	-2.90	+15.92	+2.32	+48.33	-59.77
$x \geq 7$	.0045	+.56	+.09	+.04	+.85	+2.18	+3.87	-.95	-.82	-3.60	-11.92	+51.91	-48.42	+97.38	-83.87
λ = 10															
$x \leq 2$	.0028	-1.07	-.30	-.22	+.74	+.60	+.59	-.27	-1.33	-3.54	-2.81	-23.63	+106.04	-52.51	+219.67
$x \leq 3$	.0103	-.62	-.14	-.09	+.23	+.08	-.03	+.33	+.13	-1.26	-.34	-2.79	+29.92	-31.20	+92.69
$x \leq 4$	.0293	-.39	-.07	-.04	+.12	-.02	-.15	+.30	+.53	-.35	+.68	+9.33	-1.24	-15.67	+40.14
$x \leq 5$	.0671	-.26	-.04	-.01	+.09	-.03	-.15	+.16	+.51	-.03	+.90	+14.77	-15.15	-5.84	+15.32
$x \leq 6$	.1301	-.18	-.02	-.00	+.08	-.02	-.12	+.04	+.36	+.05	+.74	+15.87	-20.89	-.44	+3.11
$x \leq 7$	.2202	-.12	-.01	+.00	+.07	-.01	-.10	-.03	+.19	+.03	+.46	+14.50	-22.17	+1.98	-2.55
$x \leq 8$	.3328	-.08	-.00	+.00	+.06	-.00	-.07	-.05	+.06	-.01	+.20	+12.00	-20.81	+2.60	-4.56
$x \leq 9$	.4579	-.06	-.00	+.00	+.05	+.00	-.05	-.05	-.03	-.04	+.00	+9.19	-17.91	+2.26	-4.53
$x \geq 11$	.4170	+.05	+.00	-.00	-.06	-.01	+.05	+.04	+.11	+.08	+.15	-9.16	+19.92	-2.17	+4.85
$x \geq 12$	.3032	+.06	+.00	-.01	-.07	-.02	+.05	+.03	+.20	+.13	+.34	-9.95	+23.97	-1.83	+4.75
$x \geq 13$	.2084	+.07	+.00	-.01	-.10	-.03	+.05	-.01	+.30	+.16	+.51	-9.97	+26.43	-.69	+2.95
$x \geq 14$	.1355	+.07	+.00	-.01	-.12	-.04	+.05	-.06	+.39	+.18	+.63	-9.04	+26.45	+1.45	-.99
$x \geq 15$	.0835	+.08	+.00	-.00	-.14	-.05	+.05	-.12	+.44	+.15	+.62	-7.02	+23.36	+4.79	-7.30
$x \geq 16$	.0487	+.08	+.00	-.00	-.15	-.05	+.06	-.17	+.45	+.07	+.43	-3.73	+16.79	+9.57	-15.89
$x \geq 17$	.0270	+.09	+.01	-.00	-.14	-.03	+.10	-.21	+.39	-.09	-.01	+1.05	+6.83	+16.07	-26.35
$x \geq 18$	.0143	+.09	+.01	+.00	-.10	+.02	+.17	-.21	+.24	-.34	-.76	+7.57	-5.95	+24.63	-37.99
$x \geq 19$	.0072	+.10	+.01	+.01	-.02	+.11	+.28	-.16	-.00	-.68	-1.89	+16.19	-20.60	+35.72	-49.98
$x \geq 20$	.0035	+.10	+.01	+.01	+.12	+.26	+.45	-.04	-.36	-1.15	-3.45	+27.38	-35.93	+49.93	-61.46

TABLE 6.3. Event $\underline{x} \le k$ or $\underline{x} \ge k+1$, exact Poisson probability and relative tail error in per cent. for some normal approximations

Event	Proba-bility	Peizer-Pratt=(5.1) ε=0	ε=.02	ε=.022	square root=(5.3) c=0	c=1/6	c=1/2	(5.11)	(4.4) γ=2/3	(4.4) γ=5/8	(4.34) c=1/2	(3.7) a=1,b=0	(3.1) c=0	(3.7) a=3/4,b=0	(3.1) c=1/2
λ = 30															
x ≤ 15	.0019	-.11	-.03	-.02	+.04	+.01	-.01	-.08	-.45	-1.03	-1.73	-19.59	+58.40	-34.50	+108.29
x ≤ 17	.0073	-.07	-.01	-.01	+.03	+.00	-.02	+.05	-.01	-.41	-.51	-6.86	+21.19	-21.18	+54.60
x ≤ 19	.0219	-.05	-.01	-.00	+.03	+.00	-.02	+.06	+.16	-.12	+.07	+1.52	+1.97	-11.30	+26.26
x ≤ 21	.0544	-.04	-.00	-.00	+.02	+.00	-.02	+.04	+.18	+.00	+.26	+6.14	-7.84	-4.74	+10.84
x ≤ 23	.1146	-.03	-.00	+.00	+.02	+.00	-.02	+.01	+.14	+.03	+.25	+7.93	-12.23	-.92	+2.63
x ≤ 25	.2084	-.02	-.00	+.00	+.01	+.00	-.02	-.00	+.08	+.03	+.17	+7.84	-13.30	+.91	-1.30
x ≤ 27	.3329	-.01	-.00	+.00	+.01	+.00	-.01	-.01	+.03	+.01	+.07	+6.69	-12.30	+1.45	-2.65
x ≤ 29	.4757	-.01	+.00	+.00	+.01	+.00	-.01	-.01	-.01	-.01	+.00	+5.10	-10.12	+1.27	-2.54
x ≥ 32	.3814	+.01	-.00	-.00	-.01	-.00	+.01	+.01	+.04	+.02	+.06	-5.68	+12.12	-1.30	+2.82
x ≥ 34	.2556	+.01	+.00	-.00	-.02	-.00	+.01	+.00	+.08	+.04	+.14	-6.23	+14.24	-.93	+2.29
x ≥ 36	.1574	+.02	+.00	-.00	-.02	-.01	+.01	-.01	+.12	+.06	+.21	-6.04	+14.79	+.22	+.17
x ≥ 38	.0890	+.02	+.00	-.00	-.03	-.01	+.01	-.02	+.15	+.05	+.22	-4.89	+13.04	+2.39	-4.00
x ≥ 40	.0463	+.02	+.00	-.00	-.03	-.01	+.01	-.04	+.16	+.02	+.15	-2.56	+8.48	+5.84	-10.45
x ≥ 42	.0221	+.02	+.00	+.00	-.02	-.00	+.02	-.04	+.13	-.06	-.05	+1.20	+.90	+10.88	-19.11
x ≥ 44	.0097	+.02	+.00	+.00	-.01	+.01	+.04	-.04	+.04	-.19	-.45	+6.70	-9.49	+17.86	-29.58
x ≥ 46	.0040	+.02	+.00	-.00	+.01	+.04	+.07	-.01	-.11	-.40	-1.09	+14.32	-22.05	+27.27	-41.18
λ = 200															
x ≤ 158	.0012	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.09	-.17	-.37	-9.98	+23.32	-15.77	+38.28
x ≤ 163	.0040	-.01	-.00	-.00	-.00	-.00	-.01	-.00	-.02	-.08	-.17	-5.17	+11.72	-10.55	+23.84
x ≤ 168	.0114	-.00	+.00	+.00	+.00	+.00	-.00	+.00	+.01	-.03	-.04	-1.65	+4.05	-6.51	+14.04
x ≤ 173	.0284	-.00	+.00	+.00	+.00	+.00	-.00	+.00	+.03	-.01	+.02	+.72	-.82	-3.57	+7.50
x ≤ 178	.0622	-.00	-.00	-.00	+.00	+.00	-.00	+.00	+.03	+.00	+.04	+2.13	-3.66	-1.56	+3.30
x ≤ 183	.1207	-.00	-.00	+.00	+.00	+.00	-.00	+.00	+.02	+.01	+.04	+2.76	-5.02	-.33	+.78
x ≤ 188	.2092	-.00	+.00	+.00	+.00	+.00	-.00	-.00	+.01	+.01	+.03	+2.83	-5.31	+.31	-.54
x ≤ 193	.3263	-.00	-.00	+.00	+.00	+.00	-.00	-.00	+.01	+.00	+.01	+2.53	-4.89	+.53	-1.03
x ≤ 198	.4624	-.00	+.00	+.00	+.00	+.00	-.00	-.00	+.00	-.00	+.00	+2.03	-4.02	+.50	-1.00
x ≥ 204	.3980	+.00	-.00	-.00	-.00	-.00	+.00	+.00	+.00	+.00	+.01	-2.21	+4.52	-.52	+1.07
x ≥ 209	.2714	+.00	+.00	-.00	-.00	-.00	+.00	+.00	+.01	+.01	+.02	-2.53	+5.30	-.43	+.92
x ≥ 214	.1696	+.00	+.00	-.00	-.00	-.00	+.00	-.00	+.02	+.01	+.03	-2.57	+5.54	-.05	+.18
x ≥ 219	.0967	+.00	-.00	-.00	-.00	-.00	+.00	-.00	+.02	+.01	+.04	-2.23	+4.96	+.75	-1.38
x ≥ 224	.0503	+.00	+.00	-.00	-.00	-.00	+.00	-.00	+.03	+.00	+.03	-1.39	+3.32	+2.07	-3.95
x ≥ 229	.0238	+.00	+.00	+.00	-.00	+.00	+.00	-.00	+.02	-.01	-.00	+.08	+.39	+4.04	-7.68
x ≥ 234	.0102	+.00	+.00	+.00	-.00	+.00	+.00	-.00	+.01	-.03	-.06	+2.29	-3.93	+6.81	-12.64
x ≥ 239	.0040	-.00	-.00	-.00	-.00	-.00	+.00	-.00	-.02	-.07	-.17	+5.40	-9.69	+10.53	-18.81

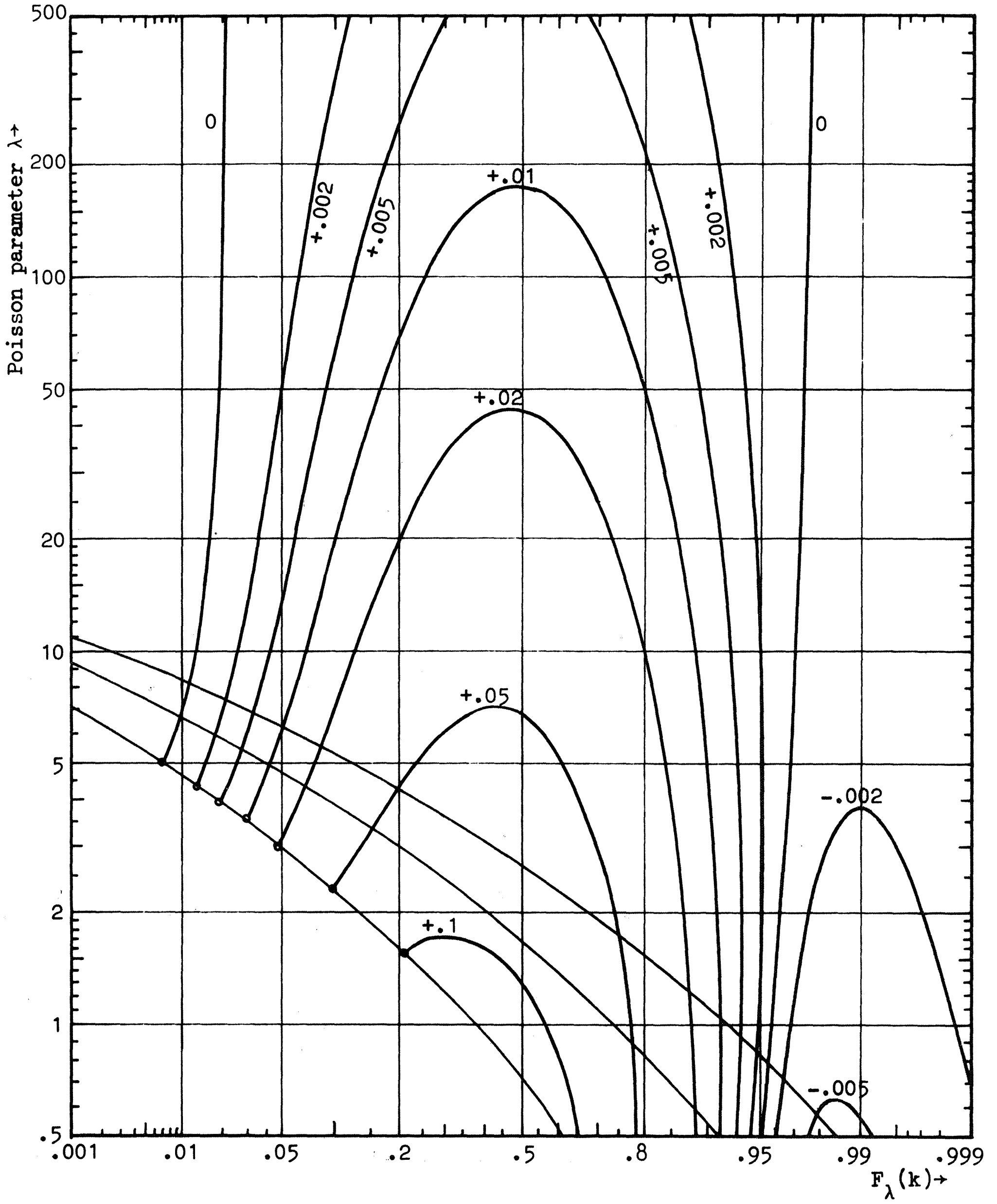

Fig. 6.1 Errors $\Phi(2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}) - F_{\lambda}(k)$ for the normal approximation (3.7) with a = 1, b = 0, to the Poisson distribution function $F_{\lambda}(k)$. Horizontal (normal probability) scale for $F_{\lambda}(k)$, vertical (logarithmic) scale for λ.

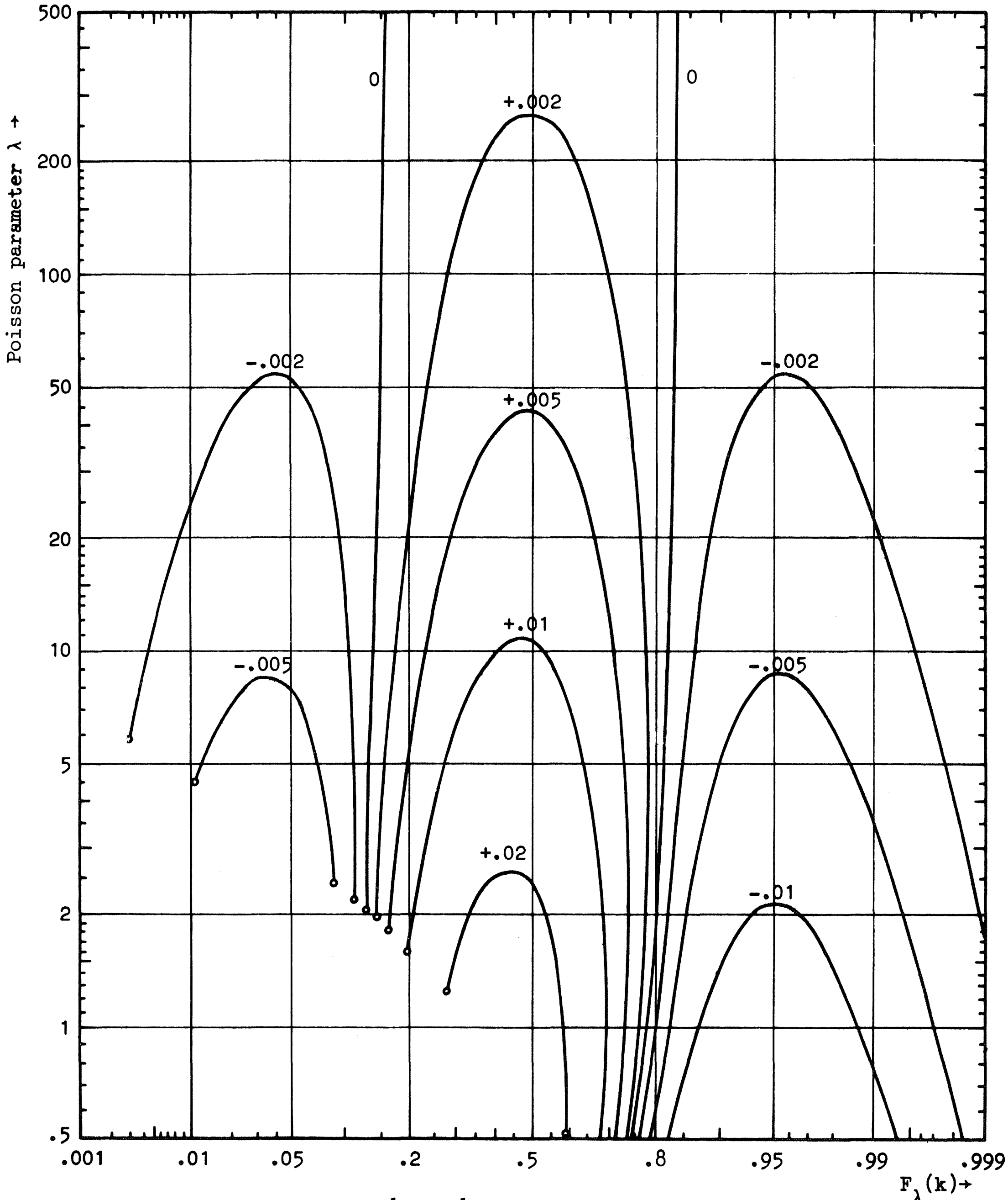

Fig. 6.2 Errors $\Phi(2(k+\frac{3}{4})^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}) - F_\lambda(k)$ for the normal approximation (3.7) with $a = \frac{3}{4}$, $b = 0$, to the Poisson distribution function $F_\lambda(k)$. Horizontal (normal probability) scale for $F_\lambda(k)$, vertical (logarithmic) scale for λ.

of P for fixed λ. Near the top of the Figures, on the other hand, the points corresponding to subsequent values of k lie so close to each other that they can hardly be distinguished.

EXAMPLE: testing H_0: $\lambda \leq 5$ against the alternative $\lambda > 5$ with $\alpha = .05$. The use of a normal approximation might lead to a wrong decision about H_0 when the error is large for $F_\lambda(k) = .95$ and $\lambda = 5$. For this pair of values one finds in Fig. 6.1 an error of +.002 and in Fig. 6.2 an error of roughly -.007. This indicates that $\Phi(2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$ could be used. The error of the more refined (5.10) does not exceed .0001, see Fig. 6.3.

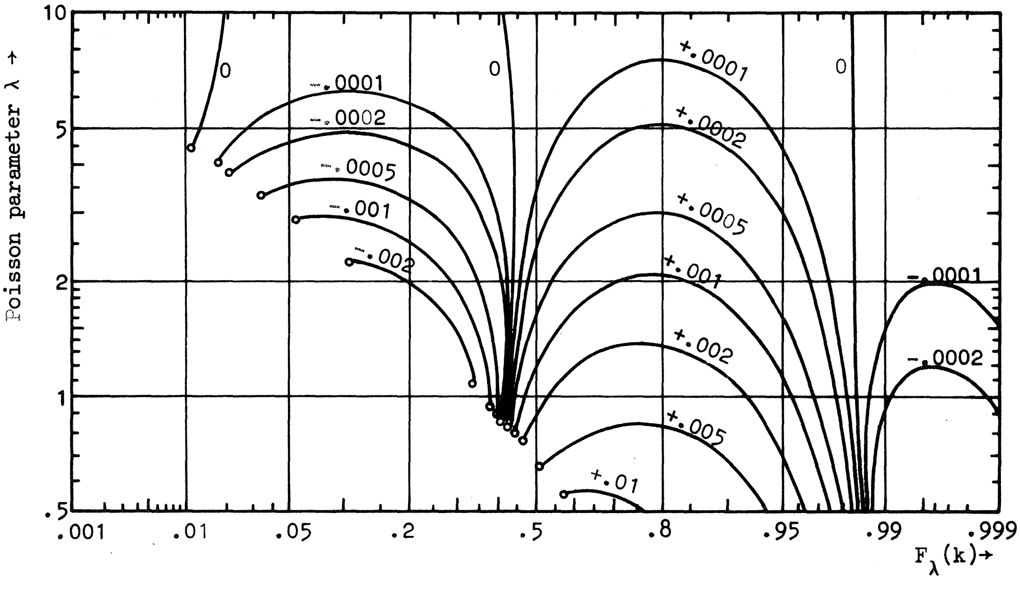

Fig. 6.3 Errors $\Phi(2\{k+(t+4)/9\}^{\frac{1}{2}} - 2\{\lambda+(t-8)/36\}^{\frac{1}{2}}) - F_\lambda(k)$, where $t = (k-\lambda+1/6)^2 \lambda^{-1}$, for the normal approximation (5.10) to the Poisson distribution function $F_\lambda(k)$. Horizontal (normal probability) scale for $F_\lambda(k)$, vertical (logarithmic) scale for λ.

CHAPTER III: NORMAL AND POISSON APPROXIMATIONS TO THE BINOMIAL DISTRIBUTION

1. NOTATION, EXACT VALUES, SUMMARY

Throughout this Chapter, the random variable $\underline{y}$ denotes the number of successes in n independent trials each having a probability p of success (n is a positive integer and $0 < p < 1$; we put $q = 1-p$ and $\sigma^2 = npq$). The distribution function of $\underline{y}$ is denoted by $G_{n,p}$:

$$(1.1) \qquad G_{n,p}(k) = P[\underline{y} \leq k] = \sum_{j=0}^{k} \binom{n}{j} p^j q^{n-j};$$

throughout the Chapter, k is an integer satisfying $0 \leq k < n$: obviously $G_{n,p}(n) = 1$.

There exist tables of $G_{n,p}(k)$ such as ORDNANCE CORPS (1952), 7 decimals, n = 1(1)150, p = .01(.01).5; WEINTRAUB (1963), 10 decimals, n = 1(1)100, p = .00001, .0001(.0001).001(.001).1; or HARVARD (1955), 5 decimals, n = 1(1)50(2)100(10)200(20)500(50)1000, p = .01(.01).5 and p = 1/16, 1/12, 1/8, 1/6, 3/16, 5/16, 1/3, 3/8, 5/12, 7/16. Because of the well known relation

$$(1.2) \qquad \begin{aligned} \sum_{j=0}^{k} \binom{n}{j} p^j q^{n-j} &= \frac{n!}{k!(n-k-1)!} \int_p^1 t^k (1-t)^{n-k-1} dt = \\ &= \frac{n!}{k!(n-k-1)!} \int_0^q s^{n-k-1} (1-s)^k ds, \end{aligned}$$

easily proved by partial integration, one may also find $G_{n,p}(k)$ from tables of the incomplete beta function, such as PEARSON (1934). But even within the ranges of any table just mentioned, approximations are often used, the tables being too bulky to be always available, and interpolation being difficult.

Suppose that $n \to \infty$, and p may vary with n. The limiting distribution is normal if and only if $npq \to \infty$, is Poisson with expectation λ if and only if $npq \to \lambda$ $(0 < \lambda < \infty)$ and is degenerate if and only if $npq \to 0$; whenever lim npq does not exist the distribution function has no limit (HEMELRIJK, 1962).

In our study of approximations to the binomial distribution function $G_{n,p}(k)$, we shall only consider $p \leq \frac{1}{2}$ (unless the contrary is explicitly stated). This is no restriction, as one can always interchange successes and failures; obviously $G_{n,p}(k) = 1 - G_{n,q}(n-k-1)$. Sections 2-5 are devoted to normal approximations (attractive because everybody possesses cumulative normal tables) and section 6 to Poisson approximations. It is sometimes difficult to decide between the two types. The just mentioned asymptotic result gives little guidance when fixed and finite values of n and p are given. Empirical determination of a Poisson-normal boundary in the (n,p)-plane is difficult for two reasons. First of all, there are many Poisson approximations, some accurate for small p only, some also for p near $\frac{1}{2}$, and many normal approximations, differing also in accuracy. Secondly, for fixed n and p a Poisson approximation may be more accurate for some probabilities and a normal one for others.

For normal approximations, section 2 gives some asymptotic results, and sections 3, 4 and 5 deal with simple, better and very accurate normal approximations respectively. As in Chapter II, these classes are defined by the asymptotic order of their error, and again this classification coincides more or less with a division as regards computational labour. Section 6 deals with approximations by a Poisson distribution, usually applied when p is small (or when p is near 1 after interchanging successes and failures). However, we shall see that a suitable choice of the Poisson parameter makes them reasonably accurate even for p near $\frac{1}{2}$. Section 7 contains numerical information about errors. A brief advice summarizing our findings is given at the end of the Chapter (p.110).

2. THE EXACT NORMAL DEVIATE

Monotonicity considerations guarantee the existence of a unique *exact normal deviate* $\xi = \xi(k,n,p)$ such that $\Phi(\xi) = G_{n,p}(k)$. This section presents a well known asymptotic expansion for ξ. It is the starting-point of our asymptotic conclusions on normal approximations, given in subsequent sections.

In all asymptotic expansions of sections 2-5 it will be tacitly understood that $\sigma^2 = npq \to \infty$ and ξ is bounded. In the literature one often finds the assumption "$n \to \infty$, p fixed" instead of "$npq \to \infty$". The latter also allows convergence of p to 0 or 1, but so slowly that $npq \to \infty$. The more restrictive assumption would suffice for our goal: starting from fixed and finite values of n and p, convergence to normality will nearly always be faster for $n \to \infty$ and p fixed than for $npq \to \infty$ but $p \to 0$ or $p \to 1$.

From "$npq \to \infty$ and ξ bounded" follows the convergence to infinity of n, np, nq, k and n-k. More precisely, one has $(k-np)/\sigma = O(1)$. Any deviate u for which $u - \xi = o(1)$ is also bounded. The corresponding values of the exact binomial distribution function, or of the approximations to it, are thus supposed to be bounded away from 0 or 1. This restriction is not serious for our purposes (cf. section II.2). The O-symbols are uniform in p; because of the restriction to bounded ξ, they can also be considered as uniform in ξ and k.

THEOREM 1

The exact normal deviate ξ defined by $\Phi(\xi) = G_{n,p}(k)$ satisfies

$$\begin{aligned}\xi = {} & u_b + \\ & + \sigma^{-1} (q-p) (- u_b^2 + 1)/6 + \\ & + \sigma^{-2} \{(5-14pq)u_b^3 + (-2+2pq)u_b\}/72 + \\ & + \sigma^{-3} (q-p) \{(-249+438pq)u_b^4 + (79-28pq)u_b^2 + 128-26pq\}/6480 + \\ & + O(\sigma^{-4}),\end{aligned} \tag{2.1}$$

where

$$u_b = (k+\tfrac{1}{2}-np)/\sigma, \text{ with } \sigma^2 = npq. \tag{2.2}$$

PROOF. Formula (2.1) is a modification of expansion (5.1) in PEIZER & PRATT (1968, II). These authors indicate very briefly a derivation along the following lines. By (1.2) the cumulative binomial probability $G_{n,p}(k)$ equals $P[\underline{\beta} \leq q]$, where $\underline{\beta}$ has a beta probability density

$s^{n-k-1}(1-s)^k\Gamma(n+1)/\{\Gamma(n-k)\Gamma(k+1)\}$. One could now apply the CORNISH-FISHER (1937) formulae to this beta variable, but for their validity it remains to show that the contributions of sixth and higher cumulants can be neglected. One may avoid this difficulty by expanding the beta density. Calculations are simplified if one considers not $(\underline{\beta} - E\underline{\beta})/\sigma(\underline{\beta})$ but

$$\underline{d} = (\underline{\beta} - \frac{n-k-1}{n-1})(n-1)^{3/2}k^{-\frac{1}{2}}(n-k-1)^{-\frac{1}{2}}. \tag{2.3}$$

The density $f(d)$ of $\underline{d}$ is expanded by using Stirling's formula for factorials and the expansion of $\log(1 \pm z)$. The result is valid for $|d| < n^{\varepsilon}$ (where ε is an arbitrary small positive number), and one integrates it after showing that the interval $(-\infty, -n^{\varepsilon})$ gives a negligible contribution. In this way VAN ZWET (1964) obtained a result which is equivalent to the first two terms of (2.1). The present author derived by the same method that

$$\begin{aligned} \xi = {} & u_b + \\ & + \sigma^{-1} (q-p) (-u_b^2 + 1)/6 + \\ & + \sigma^{-2} \{a_5u_b^5 + a_3u_b^3 + a_1u_b\} + \\ & + \sigma^{-3} \{a_6u_b^6 + a_4u_b^4 + a_2u_b^2 + a_0\} + \\ & + O(\sigma^{-4}), \end{aligned} \tag{2.4}$$

where the coefficients a_i are polynomials $\sum_j a_{ij}p^j$; the quantities a_{ij} have numerical values and do not depend on any of the variables. After some calculation one finds that $a_5 = 0$, $a_3 = (5-14p+14p^2)/72$ and $a_1 = (-2+2p-2p^2)/72$ i.e. the third term of (2.1) is correct. In principle the polynomials a_i for even i, occurring in the fourth term, can be determined in the same way, but the calculations become rather lengthy. These polynomials were therefore determined by a different method that will now be briefly sketched.

Consider the situation $n\to\infty$, p fixed, instead of $npq\to\infty$, and retain the assumption that ξ and u_b are bounded. We shall use a theorem by ESSEEN(1945, p. 61) on the distribution function of a standardized sum $\underline{y}$ of n independent identically distributed lattice random variables $\underline{y}_j$. Specializing to the present case where $P[\underline{y}_j = 0] = q$, $P[\underline{y}_j = 1] = p$, one obtains

$$
\begin{aligned}
&G_{n,p}(k+\tfrac{1}{2}) = P[(\underline{y}-np)/\sigma \leq u_b] = \\
(2.5)\qquad &= \Pi(u_b) + Q_1(z)\Pi'(u_b)n^{-\frac{1}{2}} + Q_2(z)\Pi''(u_b)n^{-1} + Q_3(z)\Pi'''(u_b)n^{-3/2} + \\
&+ o(n^{-3/2}),
\end{aligned}
$$

where Π denotes the uncorrected expansion

$$
\begin{aligned}
&\Pi(x) = \Phi(x) + \sum_{\nu=1}^{3} P_\nu(-\Phi)\, n^{-\nu/2}, \\
&P_1(-\Phi) = -\kappa_3\, \Phi^{(3)}(x)/6, \\
(2.6)\qquad &P_2(-\Phi) = -\kappa_4\, \Phi^{(4)}(x)/24 + \kappa_3^2\, \Phi^{(6)}(x)/72, \\
&P_3(-\Phi) = -\kappa_5\, \Phi^{(5)}(x)/120 - \kappa_3\kappa_4\, \Phi^{(7)}(x)/144 - \kappa_3^3\, \Phi^{(9)}(x)/1296,
\end{aligned}
$$

κ_i is the i^{th} cumulant of $(\underline{y}_j-p)p^{-\frac{1}{2}}q^{-\frac{1}{2}}$, and $Q_i(z)$ are lattice correction polynomials, given by ESSEEN, periodic modulo 1. In our case with argument $k+\frac{1}{2}$ for $G_{n,p}$ we have $z = k+[nq]-n+\frac{1}{2} \equiv \frac{1}{2}$ (mod. 1), $[\ldots]$ denoting the integer part. In that case $Q_1(z) = Q_3(z) = 0$ and $Q_2(z) = -1/24$.

One may work out (2.4) along these lines, and the first four terms of (2.1) follow by comparing the result to the Taylor expansion of

$$(2.7)\qquad G_{n,p}(k+\tfrac{1}{2}) = G_{n,p}(k) = \Phi(\xi) = \Phi(u_b + R),$$

where $R = u_b - \xi = O(n^{-\frac{1}{2}})$. This completes the derivation of (2.1) under the restricted assumption $n\to\infty$, p fixed, with $o(n^{-3/2})$ instead of $O(\sigma^{-4})$.

It is easy to remove these restrictions : the validity of (2.4) has already been established from the expansion of the beta density. The numerical values of the coefficients a_{ij} evidently remain unchanged when we pass from the assumption $n\to\infty$, p fixed to the assumption $npq\to\infty$ which includes it.

3. SIMPLE NORMAL APPROXIMATIONS

This section is devoted to normal approximations to the binomial distribution function with error $O(\sigma^{-\frac{1}{2}})$ for $\sigma \to \infty$ and bounded deviate. The three main types, here called binomial, beta and square root type, will be discussed in subsections 3a, 3b and 3c respectively. The third type, based on the well known arcsin transformation, will turn out to be generally superior. Section 3d gives results providing special accuracy near prescribed probability values.

We shall use the symbols a, b, c, d, β, γ and δ for quantities independent of ξ and σ, possibly dependent on p, which play the role of continuity corrections and similar adjustments. When an approximation has the form $\Phi(v)$ and the deviate v can be expanded as $\xi + \sigma^{-1} W(p,\xi) + O(\sigma^{-2})$, for some function W independent of σ, then Taylor expansion shows that $\Phi(v) + - G_{n,p}(k) = \sigma^{-1} W(p,\xi)\ \phi(\xi) + O(\sigma^{-2})$, as ξ is defined by $G_{n,p}(k) = \Phi(\xi)$. As long as we consider only the leading term of the error, it does not matter whether we mean the error of the deviate v or the error of the approximation $\Phi(v)$: asymptotically they differ only by the positive factor $\phi(\xi)$.

3a. SIMPLE BINOMIAL TYPE

In almost any textbook on statistics or probability theory one finds $\Phi(u_b)$ as an approximation to $G_{n,p}(k)$, where $u_b = (k+\frac{1}{2}-np)/\sigma$ and $\sigma^2 = npq$. We shall study the more general deviate

$$(3.1)\qquad (k+c-np)\{(n+d)pq+\delta\}^{-\frac{1}{2}},$$

where c, d and δ are constants or functions of p. Most authors put $\delta = d = 0$ and use the value $\frac{1}{2}$ for the continuity correction c. One finds c = q, d = 0 or 1, δ = 0, in FELLER (1945, 1957)[*)], SHANNON (1937), SHANNON & SPOERL (1942).

---

*) In the third edition (1968), this normal approximation to the binomial distribution is not mentioned.

The expansion (2.1) expresses the exact normal deviate ξ in terms of u_b and σ. Its inverted form (see also section I.2) turns out to be

$$
\begin{aligned}
(3.2)\qquad u_b &= \xi + \\
&+ \sigma^{-1}\ (1-2p)(\xi^2 - 1)/6 + \\
&+ \sigma^{-2}\ \{\xi^3(2p^2-2p-1) + \xi(-14p^2+14p-2)\}/72 + \\
&+ \sigma^{-3}\ \{\xi^4(6p^3-9p^2-9p+6) + \xi^2(14p^3-21p^2-21p+14) + \\
&- 32p^3+48p^2+48p-32\}/1620 + \\
&+ O(\sigma^{-4}).
\end{aligned}
$$

A little calculation, with introduction of $q = 1-p$ in order to make the result symmetric in p and q, leads to

$$
\begin{aligned}
&(k+c-np)\{(n+d)pq + \delta\}^{-\frac{1}{2}} = \\
&= \{u_b + (c-\tfrac{1}{2})/\sigma\}\{1 + (\delta+dpq)/\sigma^2\}^{-\frac{1}{2}} = \\
&= \xi + \sigma^{-1}\ \{(q-p)(\xi^2 - 1) + 6c-3\}/6 + \\
(3.3)\qquad &+ \sigma^{-2}\ \{\xi^3(-1-2pq) + \xi(-2-36\delta+14pq-36dpq)\}/72 + \\
&+ \sigma^{-3}\ \{\xi^4(q-p)(6+3pq) + \xi^2(q-p)(14-135\delta+7pq-135dpq) + \\
&+ (q-p)(-8+135\delta-4pq+135dpq) - (c-\tfrac{1}{2})(810\delta+810dpq)\}/1620 + \\
&+ O(\sigma^{-4}).
\end{aligned}
$$

In the *symmetric case* $p = q = \frac{1}{2}$, (3.3) becomes

$$
\begin{aligned}
&(2k+2c-n)(n+d+4\delta)^{-\frac{1}{2}} = \\
&= \xi + \sigma^{-1}\ (c-\tfrac{1}{2}) + \\
(3.4)\qquad &+ \sigma^{-2}\ \{-\ \xi^3 + \xi(1-6d-24\delta)\}/48 + \\
&+ \sigma^{-3}\ (\tfrac{1}{2}-c)(4\delta+d)/8 + \\
&+ O(\sigma^{-4}).
\end{aligned}
$$

For the choice $c = \frac{1}{2}$, the terms of order σ^{-1} and σ^{-3} vanish for all ξ simultaneously. The error of the deviate is now generally $O(\sigma^{-2})$, but it is $O(\sigma^{-4})$ for $\xi = 0$ and for $\xi^2 = 1\text{-}6d\text{-}24\delta$. Thus e.g. the choices $d = 0$, $\delta = 0$ or $d = -\frac{1}{2}$, $\delta = 0$ lead to asymptotic optimality at $\xi^2 = 1$ (tails of .16) and $\xi^2 = 4$ (tails of .023) respectively. FELLER's (1945, 1957) suggestion $d = 1$, $\delta = 0$ seems less adequate: compared to $d = \delta = 0$ it means that the leading term of the error is multiplied by $(-\xi^3 - 5\xi)/(-\xi^3 + \xi)$, and this factor is in absolute value always larger than unity, usually much larger. In section 3d we shall discuss choices of d and δ which give special accuracy for preassigned probability values α and $1 - \alpha$.

Next consider the *skew case* $p \neq q$. The leading term of the error is now $\sigma^{-1}\{(q\text{-}p)(\xi^2 - 1) + 6c - 3\}/6$, and there is no simple choice of c which makes this expression zero for all ξ simultaneously. The error is only $O(\sigma^{-2})$ for $\xi^2 = 1 + (3\text{-}6c)/(q\text{-}p)$. This means $\xi^2 = 1$ if one takes $c = \frac{1}{2}$, and $\xi^2 = 4$ if one takes $c = p$. FELLER (1945, 1957) proposes $c = q$, but then a factor $(\xi^2 + 2)$ is introduced into the leading error term which was $(\xi^2 - 1)$ for $c = \frac{1}{2}$ and $(\xi^2 - 4)$ for $c = p$. It is clear that this will make the error larger than necessary. Special choices of c providing accuracy near preassigned probability levels will be discussed in section 3d.

The choice of d or δ has in the skew case no influence on the leading term of the error. For $c = \frac{1}{2}$, $d = 1/3$, $\delta = -1/12$ the error is generally $O(\sigma^{-1})$ but $O(\sigma^{-3})$ at $\xi^2 = 1$. The same property at $\xi^2 = 4$ is obtained for $c = p$, $d = 1/6$, $\delta = -1/6$. In numerical investigations these choices provided no serious improvement when compared to $c = \frac{1}{2}$, $d = \delta = 0$ and $c = p$, $d = -\frac{1}{2}$, $\delta = 0$ respectively. Yet an improvement at $\xi^2 = 4$ or 1 was expected. A possible explanation is that for large σ the local error is already very small when the term of order σ^{-1} vanishes. For small σ the vanishing of the term of order σ^{-2} would help for probability values very close to $\Phi(\pm 1)$ or $\Phi(\pm 2)$, but such values will then be rare because of the discrete character of the binomial distribution. However, we return to this point in section 3d, where Table 3.3 gives special values of d and δ making (3.3) optimal near prescribed probability values.

When p is only slightly smaller than $\frac{1}{2}$ (we recall that $p > \frac{1}{2}$ will not be considered), the numerical values of the terms of order σ^{-1} and σ^{-2} in

(3.3) may have approximately the same magnitude. The approximation will then be accurate for small right hand tails (where the terms have opposite signs) and inaccurate for small left hand tails (where they have the same sign). In the example $n = 100$, $p = .47$, the actual probability $P[\underline{y} \geq 58] = .0177$ is estimated by the classical normal approximation ($c = \frac{1}{2}$, $d = \delta = 0$) with a relative error of $-.2$ per cent., whereas $P[\underline{y} \leq 36] = .0171$ is given with a relative error of $+3.3$ per cent.

The situation is not as simple as this example might suggest: as the two terms vary in a rather different way with ξ, they can never cancel each other for many values of ξ simultaneously. However, as long as we combine the choice of c given for a skew case, with the choice of d and δ that is good for a symmetric case, the result will also be accurate in situations which are neither symmetric nor very skew, mainly because pq is still very close to $\frac{1}{4}$ in such situations.

Summarizing our asymptotic considerations we may expect that $\Phi((k+c-np)\{(n+d)pq + \delta\}^{-\frac{1}{2}})$ is an approximation optimal for tails of .16 (i.e. $\xi = \pm 1$) if one takes $c = \frac{1}{2}$ and $d = \delta = 0$; the leading term of the error is then

$$(3.5) \qquad \sigma^{-1}\,(q-p)(\xi^2 - 1)/6 \quad \text{and} \quad \sigma^{-2}\,(-\,\xi^3 + \xi)/48$$

for $p \neq q$ and $p = q$ respectively. For optimality at $\xi = \pm 2$ (tails of .023) one may take $c = p$, $d = -\frac{1}{2}$, $\delta = 0$, for which the leading term of the error becomes

$$(3.6) \qquad \sigma^{-1}\,(q-p)(\xi^2 - 4)/6 \quad \text{and} \quad \sigma^{-2}\,(-\,\xi^3 + 4\xi)/48$$

for $p \neq q$ and $p = q$ respectively.

The validity of these conclusions for finite σ was checked in a numerical investigation including various values of n and p. Even for n as small as 20, optimality was reached for probabilities very near .16 or .023. Other values of c, d and δ were nearly always worse in this respect. FELLER's suggestion $c = q$ gave rather unsatisfactory results.

The following conclusions follow from the signs of the terms in (3.3), and from the numerical verification. For the symmetric case $p = \frac{1}{2}$, the simple binomial approximations overestimate small tails (less than the value of optimal accuracy) and underestimate larger tails. For left hand tails this remains true for $p < \frac{1}{2}$, for right hand tails the reverse holds as soon as (roughly) $\frac{1}{2}-p > n^{-\frac{1}{2}}$ (for $c = p$, $d = -\frac{1}{2}$, $\delta = 0$), or $\frac{1}{2}-p > 3n^{-\frac{1}{2}}$ (for $c = \frac{1}{2}$, $d = 0$, $\delta = 0$).

3b. SIMPLE BETA TYPE

It follows from (1.2) that $G_{n,p}(k)$ equals $P[\underline{\beta} \leq q]$, where $\underline{\beta}$ has a beta distribution with expectation $E\underline{\beta} = (n-k)(n+1)^{-1}$ and variance $\sigma^2(\underline{\beta}) = (n-k)(k+1)(n+1)^{-2}(n+2)^{-1}$. By considering $(\underline{\beta}-E\underline{\beta})/\sigma(\underline{\beta})$ which is asymptotically normal, it follows that $G_{n,p}(k)$ is approximately equal to

$$\Phi((q - \frac{n-k}{n+1})\{\frac{(n+1)^2(n+2)}{(n-k)\ (k+1)}\}^{\frac{1}{2}}) = \Phi((k+q-np)(n+2)^{\frac{1}{2}}(n-k)^{-\frac{1}{2}}(k+1)^{-\frac{1}{2}}). \tag{3.7}$$

We expand a slightly more general deviate, and find from a calculation not reproduced here that

$$\begin{aligned}(k+c-np)(n+d)^{\frac{1}{2}}(n-k-b)^{-\frac{1}{2}}(k+a)^{-\frac{1}{2}} &= \\ = \xi + \sigma^{-1}\{(q-p)(-2\xi^2-1)+6c-3\}/6 &+ \\ + \sigma^{-2}\{\xi^3(7-13pq) + \xi(18dpq-17pq+18p(b-1)-18q(a-1)) &+ \\ + 5-18c(q-p))\}/36 &+ \\ + O(\sigma^{-3}).&\end{aligned} \tag{3.8}$$

In the *symmetric case* $p = q$ one should clearly take $c = \frac{1}{2}$; the result has generally error $O(\sigma^{-2})$ and for $\xi = 0$ or $\xi^2 = (12a-12b-6d-1)/5$ this becomes $O(\sigma^{-3})$, actually even $O(\sigma^{-4})$ because the term of order σ^{-3}, not explicitly given in (3.8), vanishes identically in ξ for $p = q$. In the *skew case* the error is generally of order σ^{-1}, but for $\xi^2 = -\frac{1}{2} + (3c-1\frac{1}{2})/(q-p)$ it is $O(\sigma^{-2})$. This means $c = q$ for optimality near $\xi^2 = 1$ and $c = 2 - 3p$ for optimality near $\xi^2 = 4$, but anyhow the leading term of the error is twice as large, with opposite sign, when compared to the simple binomial type with $c = \frac{1}{2}$, $d = \delta = 0$ or $c = p$, $d = -\frac{1}{2}, \delta = 0$ respectively. More generally, for any given simple beta type deviate one can find a simple binomial type one which is asymptotically twice as accurate for $p \neq \frac{1}{2}$.

Numerical investigations confirm the inferiority of the beta type. From theoretical and numerical considerations it follows that the choices $c = q$, $d = -1/3$, $a = 2/3$, $b = 1/3$ produce accurate results near $\xi = \pm 1$, whereas for $\xi = \pm 2$ one may put $c = 2-3p$, $d = -13/6$, $a = 5/6$, $b = 1/6$. As we shall encounter in section 3c approximations which are much better and less cumbersome, there is no reason for further consideration of beta type deviates.

3c. ARCSIN AND SIMPLE SQUARE ROOT TYPE

It is sometimes desired to use analysis of variance in order to test the hypothesis that several binomial success probabilities are equal. The observed numbers of successes have approximately a normal distribution when the number of experiments in each run is sufficiently large. However, the variance of the binomial variable depends functionally on the mean, which implies that the observations do not have constant variance under the alternative hypothesis that their means are unequal.

One could try to replace the binomial variable $\underline{y}$ by some transformed variable $\Psi(\underline{y})$ with variance independent of its mean. However, CURTISS (1943) has shown that such a transformation does not exist (except the trivial one Ψ = constant). We shall now sketch the derivation of a transformation Ψ for which the functional dependence of variance on mean vanishes asymptotically for $n \to \infty$ and constant p.

Let $\underline{f}$ denote the fraction of successes $\underline{y}/n$. As $\underline{f}$-p is nearly always small when n is large, we may roughly say that

$$(3.9)\qquad \begin{aligned} &\Psi(\underline{f}) \simeq \Psi(p) + (\underline{f}-p)\Psi'(p); \\ &E\Psi(\underline{f}) \simeq E\{\Psi(p) + (\underline{f}-p)\Psi'(p)\} = \Psi(p); \\ &\sigma^2(\Psi(\underline{f})) \simeq E\{\Psi(\underline{f}) - \Psi(p)\}^2 \simeq E(\underline{f}-p)^2\{\Psi'(p)\}^2; \\ &\sigma(\Psi(\underline{f})) \simeq \Psi'(p)\sigma(\underline{f}) = \Psi'(p)\{p(1-p)/n\}^{\frac{1}{2}}. \end{aligned}$$

If this standard deviation must be almost independent of p and n, one has roughly $\Psi'(p) = Kn^{\frac{1}{2}}p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$, which means $\Psi(p) = Kn^{\frac{1}{2}}\arcsin p^{\frac{1}{2}} + M$ (K and M are arbitrary constants). Thus the heuristic approach indicates that the transformation $\Psi_1(\underline{f}) = n^{\frac{1}{2}}\arcsin \underline{f}^{\frac{1}{2}}$ might approximately stabilize the variance of the fraction of successes $\underline{f} = \underline{y}/n$. Now two theorems by CURTISS (1943) furnish a neat proof that

$$(3.10) \qquad 2n^{\frac{1}{2}}\{\arcsin \underline{f}^{\frac{1}{2}} - \arcsin p^{\frac{1}{2}}\} \qquad \text{(in radians)}$$

has asymptotically ($n\to\infty$, p fixed) a standard normal distribution, and a variance tending to unity. The arguments of CURTISS can be extended to the case where $npq \to \infty$ but $p \to 0$; even then the transformation maintains asymptotic normality and makes the variance asymptotically independent of the mean. Tables of $2 \arcsin p^{\frac{1}{2}}$ (HALD, 1952a; DE JONGE, 1960) facilitate its application.

ANSCOMBE (1948) finds that

$$(3.11) \qquad \Psi_2(\underline{y}) = (n+\delta)^{\frac{1}{2}} \arcsin \{(\underline{y}+c)^{\frac{1}{2}}/(n+\gamma)^{\frac{1}{2}}\},$$

again expressed in radians, has for $n \to \infty$ and p fixed a variance

$$(3.12) \qquad \sigma^2(\Psi_2(\underline{y})) = \tfrac{1}{4} \{1 + \frac{2\delta-1}{2n} + \frac{3-8c}{8np} + \frac{3+8c-8\gamma}{8nq} + O(n^{-2})\}.$$

He proposes the choices $\delta = \frac{1}{2}$, $c = 3/8$, $\gamma = \frac{3}{4}$, for which one has $\sigma^2(\Psi_2(\underline{y})) = \frac{1}{4} + O(n^{-2})$. According to FREEMAN & TUKEY (1950),

$$(3.13) \qquad \Psi_3(\underline{y}) = (n+\tfrac{1}{2})^{\frac{1}{2}} \left[\arcsin \{\underline{y}/(n+1)\}^{\frac{1}{2}} + \arcsin \{(\underline{y}+1)/(n+1)\}^{\frac{1}{2}}\right]$$

gives excellent variance stabilization; see also LAUBSCHER (1961).

Our purpose is not variance stabilization but normal approximation to the binomial distribution function $G_{n,p}(k)$. For arbitrary constants β, γ, δ, we find, from a series expansion using the derivatives of $\arcsin p^{\frac{1}{2}}$, that

(3.14)
$$2(n+\delta)^{\frac{1}{2}}\left[\arcsin\{(k+\tfrac{1}{2}+\beta)/(n+\gamma)\}^{\frac{1}{2}} - \arcsin p^{\frac{1}{2}}\right] =$$
$$= \xi + \sigma^{-1}\{(q-p)\,(-\xi^2 - 2) + 12\beta - 12\gamma p\}/12 +$$
$$+ \sigma^{-2}\{\xi^3(1-pq) + \xi(2-18\beta+36\beta p-18\gamma p-5pq+18\delta pq)\}/36 +$$
$$+ 0(\sigma^{-3}).$$

The deviate in the left hand side of (3.14) is used with $\beta = \gamma = \delta = 0$ by HALD (1952) and RAFF (1955, 1956), with $\beta = \frac{1}{2}$, $\gamma = \delta = 1$ by FREEMAN & TUKEY (1950). ANSCOMBE's best transformation (3.11) would mean $\beta = 3/8$, $\gamma = \frac{3}{4}$, $\delta = \frac{1}{2}$ when a continuity correction of $\frac{1}{2}$ is used.

If one takes in (3.14) $\beta = q\gamma$, $\beta = \frac{1}{2}\gamma$ or $\beta = p\gamma$, the leading term of the error for any $p \neq q$ becomes proportional to $(q-p)$. That is the only situation in which simple choices of β, γ and δ lead to asymptotic optimality at fixed values of ξ. The leading term contains a factor $(-\xi^2-2+12\gamma)$ for $\beta = q\gamma$, a factor $(-\xi^2-2+6\gamma)$ for $\beta = \frac{1}{2}\gamma$, a factor $(-\xi^2-2)$ for $\beta = p\gamma$. It follows that $\gamma = \frac{1}{2}$, $\beta = \frac{1}{4}$ or $\gamma = \frac{1}{4}$, $\beta = q/4$ gives asymptotic optimality at $\xi^2 = 1$ (tails of .16); similarly $\gamma = 1$, $\beta = \frac{1}{2}$ or $\gamma = \frac{1}{2}$, $\beta = \frac{1}{2}q$ will do well near $\xi^2 = 4$ (tails of .023). For $\beta = p\gamma$ the leading term does not vanish for any ξ (unless $p = q$).

The value of δ does not affect the leading term; we may choose it in such a way that the next term vanishes too for the relevant value of ξ^2. In Table 3.1 such values are denoted by δ_p. The value $\delta_{\frac{1}{2}}$, also given in the Table, is found by substituting $p = \frac{1}{2}$. As pq varies between .21 and .25 for $.3 \leq p \leq .7$, we expect that $\delta_{\frac{1}{2}}$ will not be much worse than δ_p for such values of p.

When β, γ and δ_p are chosen according to either of the first two lines of Table 3.1, (3.14) gives a deviate

(3.15)
$$\xi + \sigma^{-1}(q-p)(-\xi^2+1)/12 + \sigma^{-2}(1-pq)(\xi^3-\xi)/36 + 0(\sigma^{-3});$$

for the last two lines, it gives

(3.16)
$$\xi + \sigma^{-1}(q-p)(-\xi^2+4)/12 + \sigma^{-2}(1-pq)(\xi^3-4\xi)/36 + 0(\sigma^{-3}).$$

TABLE 3.1

Values of β, γ, δ_p and $\delta_{\frac{1}{2}}$ (see text) such that (3.14) gives a good approximation near the stated values of ξ^2

	β	γ	δ_p	$\delta_{\frac{1}{2}}$
$\xi^2 = 1$	1/4	1/2	1/3 + 1/(12pq)	2/3
	q/4	1/4	- 1/6 + 1/(12pq)	1/6
$\xi^2 = 4$	1/2	1	1/2 + 1/(6pq)	7/6
	q/2	1/2	- 1/2 + 1/(6pq)	1/6

For $p = q = \frac{1}{2}$, the same expansions hold also for any other triplet (β,γ,δ) with $\beta = \frac{1}{2}\gamma$ and $2\gamma-\delta = 1/3$ (case $\xi^2 = 1$) or $\beta = \frac{1}{2}\gamma$ and $2\gamma-\delta = 5/6$ (case $\xi^2 = 4$). This follows when $p = q = \frac{1}{2}$ is substituted into (3.14). As the $O(\sigma^{-3})$ term, not explicitly given there, vanishes when $p = q$ and $\beta = \frac{1}{2}\gamma$, the error is in that case generally $O(\sigma^{-2})$ and locally $O(\sigma^{-4})$ for the special values of ξ.

The simple binomial type deviate (3.3) has for $c = \frac{1}{2}$ and $d = \delta = 0$ an error with leading term (3.5), and for $c = p$, $\delta = 0$, $d = -\frac{1}{2}$ with leading term (3.6). By comparison with (3.15) and (3.16), it follows that the arcsin deviate (3.14) with constants as in Table 3.1 is asymptotically twice as accurate (BORGES, 1970) as (3.3) for $p \neq q$, and just as accurate for $p = q$; the sign of the error is reversed. This conclusion continues to hold for deviates optimal at other values than $\xi^2 = 1$ or 4.

HALD (1952 , section 21.6) gives (3.14) with $\beta = \gamma = \delta = 0$ and compares it to (3.3) with $c = \frac{1}{2}$, $d = \delta = 0$. He observes: "Systematic examinations of the accuracy of this approximation formula have not been made, but the scattered experience we have indicates that the two approximation formulas usually lead to deviations from the exact values of the same order of magnitude, but with opposite signs", and suggests to take the average of the two values. The series expansions given above show that he is right as regards the signs, but asymptotically the errors will not cancel.

The possibility to take $(p+V)^{\frac{1}{2}}$ instead of $p^{\frac{1}{2}}$ in (3.14) gives no improvement unless a complicated expression for V is used. Therefore this idea was discarded.

The arcsin type deviate (3.14) may be more accurate than the simple binomial or beta types, it is also more cumbersome. However, it has a modified form in which the arcsin transformation is avoided, and which is usually at least as accurate. This modification, here called *simple square root type*, is related to binomial probability paper (MOSTELLER & TUKEY, 1949). Following the exposition in this reference, we shall give a geometrical interpretation of (3.14), for $\beta = \frac{1}{2}$, $\gamma = \delta = 1$. Plot the points $B = (n-k,k+1)$ and $F = (q,p)$ on the binomial graph paper, which is graduated with a square root scale on both axes (Fig. 3.1). Let e.g. $|BD|$ denote the length, in linear scale, of the line segment joining the points B and D. The Figure shows that $|BD| = (k+1)^{\frac{1}{2}}$, $|OD| = (n-k)^{\frac{1}{2}}$, and $|OB| = (n+1)^{\frac{1}{2}}$ by Pythagoras' theorem. Thus $\arcsin\{(k+1)^{\frac{1}{2}}/(n+1)^{\frac{1}{2}}\} = \arcsin\{|BD|/|OB|\} = \psi$. By the same reasoning, $|OE| = q^{\frac{1}{2}}$, $|EF| = p^{\frac{1}{2}}$, $|OF| = 1$ and $\arcsin p^{\frac{1}{2}} = \psi_0$.

It follows that, expressed in radians,

$$(3.17) \qquad 2(n+1)^{\frac{1}{2}}\ \{\arcsin\left(\frac{k+1}{n+1}\right)^{\frac{1}{2}} - \arcsin p^{\frac{1}{2}}\} = 2|OB|(\psi-\psi_0) = \pm\ 2|\text{arc BC}|,$$

where the + sign is valid if and only if $(k+1)/(n+1) > p$, as is the case in Fig. 3.1. Thus the deviate (3.14) can be found by measuring a certain arc in a graph.

For $n \to \infty$ and p fixed, the observed fraction of successes tends in probability to p. This means that $|\psi-\psi_0|$ will be small unless $G_{n,p}(k)$ is very close to 0 or 1. If the angle is small, it does not make much difference if one replaces arc BC by the line segment BA perpendicular to OF. Thus the distance $|AB|$ is almost proportional to a standard normal deviate, and their ratio depends only on the scale of the graph. This idea underlies the numerous applications of binomial probability paper described by MOSTELLER & TUKEY (1949).

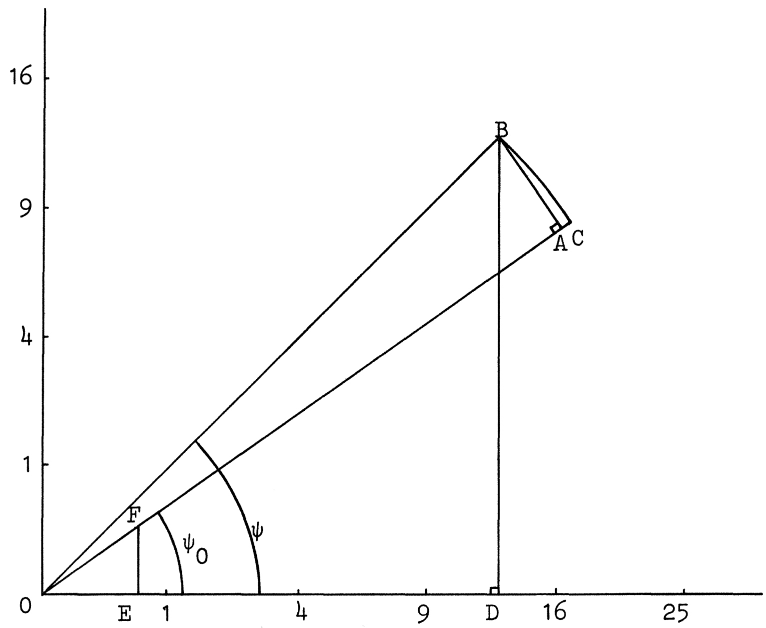

Fig. 3.1. Binomial probability paper showing the arcsin deviate 2 arc BC and the square root deviate 2AB (see text).

See also a monograph, and a series of papers in Rep. Stat. Appl. Res. of the Union of Japanese Scientists and Engineers (1951, 1952), by MASUYAMA. This very convenient graph paper is not sufficiently popular to be generally available. Therefore we now proceed to its algebraic counterpart. Obviously one has

$$
\begin{aligned}
(3.18)\qquad & \pm\, 2|BA| = 2|OB| \sin(\psi-\psi_0) = \\
&= 2|OB| \{\sin\psi \cos\psi_0 - \cos\psi \sin\psi_0\} = \\
&= 2(n+1)^{\frac{1}{2}}\{(k+1)^{\frac{1}{2}}(n+1)^{-\frac{1}{2}}q^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(n+1)^{-\frac{1}{2}}p^{\frac{1}{2}}\} = \\
&= 2(k+1)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k)^{\frac{1}{2}}p^{\frac{1}{2}}.
\end{aligned}
$$

If one starts from (3.14) in its general form, and puts $\delta = \gamma$ in order to obtain cancellation of $(n+\delta)^{\frac{1}{2}}$ and $(n+\gamma)^{-\frac{1}{2}}$, the result is $2(k+\frac{1}{2}+\beta)^{\frac{1}{2}}q^{\frac{1}{2}}$ + $-\ 2(n-k-\frac{1}{2}-\beta+\gamma)^{\frac{1}{2}}p^{\frac{1}{2}}$. It is convenient to introduce new constants $a = \frac{1}{2}+\beta$ and $b = -\frac{1}{2}-\beta+\gamma$. Then a series expansion derived from (3.3) gives

$$
\begin{aligned}
&2(k+a)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+b)^{\frac{1}{2}}p^{\frac{1}{2}} = \xi + \\
&+ \sigma^{-1}\ \{(q-p)(-\ \xi^2 - 2) + 12aq-12bp-6\}/12 + \\
(3.19)\qquad &+ \sigma^{-2}\ \{\xi^3(2-5pq) + \xi(22-46p+10p^2-36bp^2-36aq^2)\}/72 + \\
&+ O(\sigma^{-3}).
\end{aligned}
$$

The special case $a = 1$, $b = 0$ is used by MOSTELLER & TUKEY (1949) and FREEMAN & TUKEY (1950), who call it the *chordal transform.*

We shall consider pairs (a,b) with $a-b = 1$: in that case the term of order σ^{-1} is proportional to $(q-p)$, and this leads to simple values of a and b providing asymptotic optimality at fixed values of ξ. For $a = b+1$, (3.19) becomes

$$
\begin{aligned}
&2(k+1+b)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+b)^{\frac{1}{2}}p^{\frac{1}{2}} = \xi + \\
&+ \sigma^{-1}\ (q-p)(-\ \xi^2 + 4+12b)/12 + \\
(3.20)\qquad &+ \sigma^{-2}\ \{\xi^3(2-5pq) + \xi(-14-36b+26pq+72bpq)\}/72 + \\
&+ O(\sigma^{-3}).
\end{aligned}
$$

In the *skew case* $p \neq q$, optimality near $\xi^2 = 1$ is provided by the choice $b = -\frac{1}{4}$, for which the expansion in (3.20) becomes

$$
\begin{aligned}
&\xi + \sigma^{-1}(q-p)(-\ \xi^2 + 1)/12 + \\
(3.21)\qquad &+ \sigma^{-2}\ \{\xi^3(2-5pq) + \xi(-5+8pq)\}/72 + O(\sigma^{-3}).
\end{aligned}
$$

For optimality near $\xi^2 = 4$ the best choice is $b = 0$, as proposed by FREEMAN & TUKEY (1950); this gives (3.21) with $(-\ \xi^2 + 4)$ instead of $(-\ \xi^2 + 1)$ and $(-14+26pq)$ instead of $(-5+8pq)$.

As could be expected from the geometric derivation of the square root deviate, the terms of order σ^{-1} of (3.15) and (3.21) coincide. The term of order σ^{-2} will generally be smaller for (3.15), and certainly so near $\xi^2 = 1$. Similar conclusions can be drawn if (3.16) is compared to the square root deviate with $b = 0$. However, equal first term of the error and smaller second term do not guarantee that the arcsin deviate is always more accurate: when the first and second terms have opposite signs, a larger second term may give a better compensation of the first one.

The error of the simple square root type is for any $p \neq \frac{1}{2}$ asymptotically $(-\frac{1}{2})$ times the error of the corresponding version of the simple binomial type, because this property was already derived for the arcsin type.

In the *symmetric case* $p = q = \frac{1}{2}$, it is obviously advantageous to take $a = b+1$. The error now becomes

$$(3.22) \qquad \sigma^{-2}(\xi^3 - 10\xi - 24b\xi)/96 + O(\sigma^{-4}),$$

because the term of order σ^{-3}, not explicitly given in (3.19) or (3.20), vanishes. The approximation is optimal near $\xi = 0$ and $\xi^2 = 10 + 24b$. This means $b = -3/8$ if one wants accuracy near $\xi^2 = 1$, and $b = -\frac{1}{4}$ near $\xi^2 = 4$. The error for any other ξ is twice as small as it was for the arcsin or simple binomial type, see (3.5), (3.6), (3.15) and (3.16). This conclusion continues to hold for deviates optimal at other values than $\xi^2 = 1$ or $\xi^2 = 4$.

Numerical investigation shows that most of the asymptotic conclusions remain correct for finite values of n. The square root deviate (3.20) gives usually the best results, and its calculation is rather simple. It is optimal for tails of roughly .16 if one takes $b = -\frac{1}{4}$ for skew cases and $b = -3/8$ for symmetric cases, and optimal for tails of roughly .023 with $b = 0$ (skew) and $b = -\frac{1}{4}$ (symmetric). As could be expected from the expansions, the actual location of the optimum for finite n varies more for the simple square root type than for the arcsin or simple binomial types.

Numerical evidence shows that for roughly $|q-p| < n^{-\frac{1}{2}}$ the values of b corresponding to a symmetric case give better results than the values for a skew case. As was stated in section 3a, the terms of order σ^{-1} and σ^{-2} exhibit a rather different functional dependence on n and p. This makes it difficult to establish more than a footrule for a boundary between skew cases and (almost) symmetric cases.

Even when a table of 2 arcsin $p^{\frac{1}{2}}$ is available, the arcsin type offers hardly any advantage. In the symmetric case the square root type is found to be clearly more accurate, and for $p < q$ it has the property of diminishing the error for left hand tails (where it was rather large for arcsin type) at cost of a somewhat larger error for right hand tails (where it was small).

In section 7 numerical values of errors will be given. It should be emphasized that the error of all simple approximations begins with a term proportional to $(q-p)/\sigma$. This means that for constant n the errors increase rapidly if the distribution becomes more and more skew. The classical approximation $\Phi((k+\frac{1}{2}-np)/\sigma)$ estimates for n = 100 and p = .5 the value $P[\underline{y} \leq 40] = .0284$ with a relative error of +1.0 per cent. (it gives the answer .0287). For n = 100 and p = .45, the approximated value is .0281 for $P[\underline{y} \leq 35] = .0272$, the relative error is now + 3.2 per cent. For n = 100 and p = .2, however, $P[\underline{y} \leq 12] = .0253$ is approximated with a relative error of + 20.0 per cent., i.e. the approximation gives .0304.

In section 4 we shall study approximations which have an error of order σ^{-2} for both skew and symmetric cases.

3d. ACCURATE APPROXIMATION NEAR PREASSIGNED VALUES

Sometimes an approximation to cumulative binomial probabilities is desired to be accurate near the prescribed values α and/or $1-\alpha$ of the distribution function, whereas it may be rather rough elsewhere. This may happen when a two-sided hypothesis about the success probability p is tested at significance level 2α, or a one-sided hypothesis at level α.

As before, ξ_α denotes the standard normal upper α fractile , i.e. $\Phi(\xi_\alpha) = 1-\alpha$. We shall suppose that $\alpha < .5$, and thus $\xi_\alpha > 0$.

For $p \neq q$ it follows from (3.3) that for the simple binomial type the leading term of the error vanishes at $\xi = \pm\xi_\alpha$ if one chooses $c = \frac{1}{2} + (q-p)(1-\xi_\alpha^2)/6$. Values of c for various p and α are given in Table 3.2.

TABLE 3.2

Values $c = c(p,\alpha) = \frac{1}{2} + (q-p)(1-\xi_\alpha^2)/6$, such that $\Phi((k+c-np)/\sigma)$ is an approximation to $G_{n,p}(k)$ especially accurate for probabilities near α and $1-\alpha$. For $p = .5$ take $c = \frac{1}{2}$ for all α; for $p > .5$ take $1 - c(1-p,\alpha)$.

α	.1	.05	.025	.01	.005
ξ_α^2	1.642	2.706	3.841	5.412	6.635
$c(p,\alpha)$	.393+.214p	.216+.569p	.026+.947p	−.235+1.471p	−.439+1.878p
$c(.45,\alpha)$	.489	.472	.453	.426	.406
$c(.4,\alpha)$	.479	.443	.405	.353	.312
$c(.3,\alpha)$	.457	.386	.311	.206	.124
$c(.2,\alpha)$	.436	.329	.216	.059	−.063
$c(.1,\alpha)$	.414	.273	.121	−.088	−.251
$c(.05,\alpha)$	.404	.244	.074	−.162	−.345

TABLE 3.3

Values of $d = d(\alpha) = (-\xi_\alpha^2+7)/18$ and $\delta = \delta(\alpha) = (-\xi_\alpha^2-2)/36$, such that $\Phi((k+c-np)\{(n+d)pq+\delta\}^{-\frac{1}{2}})$, with c given by Table 3.2, is an approximation to $G_{n,p}(k)$ especially accurate for probabilities near α and $1-\alpha$. For $p = q = \frac{1}{2}$ one may use (3.4) with $c = \frac{1}{2}$ and $d + 4\delta = (1-\xi_\alpha^2)/6$ given below.

α	.1	.05	.025	.01	.005
$d(\alpha)$	+.298	+.239	+.175	+.088	+.020
$\delta(\alpha)$	−.101	−.131	−.162	−.206	−.240
$d+4\delta$	−.107	−.284	−.474	−.735	−.939

The denominator of the simple binomial deviate can be $\sigma = (npq)^{\frac{1}{2}}$, but some improvement may be expected if one uses $\{(n+d)pq + \delta\}^{\frac{1}{2}}$, with d and δ such that the term of order σ^{-2} in (3.3) vanishes too at $\xi = \pm\xi_\alpha$. This is the case for $d = (-\xi_\alpha^2+7)/18$ and $\delta = (-\xi_\alpha^2-2)/36$, which are given for various α in Table 3.3.

Similar tables could be given for constants making the simple beta or arcsin type optimal near α and $1-\alpha$. As this would lead to laborious approximations, we pass directly to the simple square root type. Here (3.20) implies that

$$\Phi(2(k+1+b)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+b)^{\frac{1}{2}}p^{\frac{1}{2}}) \tag{3.23}$$

has an error $O(\sigma^{-2})$ at $\xi = \pm\xi_\alpha$ if one puts $b = b(\alpha) = (\xi_\alpha^2-4)/12$, a value valid for all p. Going back to (3.19), one can determine a and b such that the terms of order σ^{-1} and σ^{-2} vanish both at $\xi = \pm\xi_\alpha$. Some calculation shows that this is achieved by the approximation

$$\Phi(2\{(k+1)q + B(p,\alpha)\}^{\frac{1}{2}} - 2\{(n-k)p + B(q,\alpha)\}^{\frac{1}{2}}), \tag{3.24}$$

where

$$B(p,\alpha) = (-p^2-7q)/18 + (2q-p^2)\xi_\alpha^2/36. \tag{3.25}$$

Table 3.4 gives $b(\alpha)$ and $B(p,\alpha)$ for various α and p.

Let us now compare the four proposed approximations. It is obvious that (3.23), with $b(\alpha) = (\xi_\alpha^2-4)/12$ given in the second line of Table 3.4, is much easier than the other three. Numerical investigation shows that these other three may be somewhat more accurate for probabilities of exactly α or $1-\alpha$. However, for $\xi \neq \xi_\alpha$ the error of (3.23) and (3.24) is $\sigma^{-1}(q-p)(-\xi^2 + \xi_\alpha^2)/12 + O(\sigma^{-2})$, whereas it is $\sigma^{-1}(q-p)(\xi^2 - \xi_\alpha^2)/6 + O(\sigma^{-2})$ for binomial type with c as given in Table 3.2. This makes it clear why the square root type (3.23) and (3.24) are generally better unless the unknown probability lies very close to α or $1-\alpha$. This will rarely happen because of the discrete nature of the binomial distribution.

TABLE 3.4

Values $b = b(\alpha) = (\xi_\alpha^2-4)/12$ and $B(p,\alpha)$, see (3.25), such that (3.23) and (3.24) are especially accurate for probabilities near α and $1-\alpha$.

α	.1	.05	.025	.01	.005
$b(\alpha)$	-.196	-.108	-.013	+.118	+.220
$B(.05,\alpha)$	-.283	-.227	-.167	-.084	-.020
$B(.1,\alpha)$	-.269	-.216	-.160	-.081	-.021
$B(.2,\alpha)$	-.242	-.196	-.147	-.079	-.026
$B(.3,\alpha)$	-.217	-.179	-.137	-.080	-.036
$B(.4,\alpha)$	-.195	-.164	-.131	-.086	-.051
$B(.45,\alpha)$	-.184	-.158	-.129	-.090	-.060
$B(.5,\alpha)$	-.174	-.152	-.128	-.096	-.070
$B(.55,\alpha)$	-.165	-.147	-.128	-.102	-.082
$B(.6,\alpha)$	-.155	-.142	-.129	-.109	-.094
$B(.7,\alpha)$	-.139	-.136	-.132	-.127	-.124
$B(.8,\alpha)$	-.124	-.131	-.139	-.149	-.158
$B(.9,\alpha)$	-.112	-.130	-.149	-.176	-.196
$B(.95,\alpha)$	-.106	-.130	-.155	-.190	-.217

So far we have concentrated our attention on the case $p \neq q$. For $p = q = \frac{1}{2}$ one may use the simple binomial deviate in the form (3.4) with $c = \frac{1}{2}$ and $d + 4\delta = (1-\xi_\alpha^2)/6$ (last line of Table 3.3), or the square root type (3.24) where now $B(.5,\alpha) = (\xi_\alpha^2-10)/48$, see Table 3.4. However, for $p = q$ it will usually not be much extra trouble to use an approximation which is $O(\sigma^{-4})$ for all ξ, not only for $\xi = \pm\xi_\alpha$. We refer to section 5a, where such approximations are discussed.

4. BETTER NORMAL APPROXIMATIONS

So far we have discussed simple deviates, for which the error is $O(\sigma^{-1})$ for $p \neq q$, and $O(\sigma^{-2})$ for $p = q$. The present section deals with approximations with error $O(\sigma^{-2})$ for all p. Among these are the CAMP-PAULSON approximation (section 4a), the BORGES approximation (section 4b) and improved versions of simple binomial and square root type approximations (section 4c). Section 4d gives a comparison of their series expansions and some information about their errors.

4a. CAMP-PAULSON APPROXIMATION

There exist well known relations between SNEDECOR's F-distribution and the beta distribution, and also between the beta and binomial distribution functions, cf.(1.2). They imply that

(4.1) $$G_{n,p}(k) = P\left[\underline{F} \leq F\right],$$

where

(4.2) $$F = \frac{(k+1)q}{(n-k)p},$$

and $\underline{F}$ has $\nu_1 = 2n-2k$ degrees of freedom in the numerator and $\nu_2 = 2k+2$ degrees of freedom in the denominator.

If $\underline{w}_i$ (i=1,2) are two normally distributed random variables with expectations μ_i, variances σ_i^2 and correlation coefficient ρ, then $(\underline{w}_1\mu_2/\underline{w}_2 - \mu_1)/\{\underline{w}_1^2\sigma_2^2/\underline{w}_2^2 - 2\rho\underline{w}_1\sigma_1\sigma_2/\underline{w}_2 + \sigma_1^2\}^{\frac{1}{2}}$ has approximately a standard normal distribution provided that $P(\underline{w}_1 < 0)$ is small (GEARY, 1930; FIELLER, 1932).

Now SNEDECOR's $\underline{F}$ is the ratio of the independent random variables $\underline{w}_1 = \chi^2_{\nu_1}/\nu_1$ and $\underline{w}_2 = \chi^2_{\nu_2}/\nu_2$. For large ν_i, $\underline{w}_i$ has approximately a normal distribution with expectation 1 and variance $2/\nu_i$. Thus the FIELLER-GEARY result implies that (4.1) is approximately equal to

$$(4.3) \qquad \Phi(\{F-1\}/\{2F^2/\nu_2 + 2/\nu_1\}^{\frac{1}{2}}) = = \Phi(\{k+q-np\}\{(n-k)p^2 + (k+1)q^2\}^{-\frac{1}{2}}).$$

The normal deviate occurring in the second half of (4.3), obtained by expressing F, ν_1 and ν_2 in terms of k, n, p, q, has the asymptotic expansion

$$(4.4) \qquad \xi + \sigma^{-1}(q-p)(-\xi^2+1)/6 + \sigma^{-2}(7\xi^3-31pq\xi^3-13\xi+37pq\xi)/36 + O(\sigma^{-3}).$$

This means for $p \neq q$ that it will be asymptotically just as accurate as a simple beta type deviate (3.8) with c = q. For p = q, the deviate in (4.3) becomes $\{k+\frac{1}{2}-np\}\{(n+1)/4\}^{-\frac{1}{2}}$, i.e. (3.4) with d = 1 and $\delta = 0$. Thus (4.3) is more laborious, and not more accurate, than the simple binomial type, and is not considered any further.

The FIELLER-GEARY result can also be applied to $\underline{F}^{\frac{1}{2}}$, which is the ratio of two independent random variables $(\underline{\chi}^2_{\nu_i}/\nu_i)^{\frac{1}{2}}$ that are asymptotically ($\nu_i \to \infty$) normal with expectation 1 and variance $(2\nu_i)^{-1}$. It follows now that (4.1) is approximately equal to

$$(4.5) \qquad \Phi(\{F^{\frac{1}{2}}-1\}/\{F/(2\nu_2) + 1/(2\nu_1)\}^{\frac{1}{2}}) = \Phi(2(k+1)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k)^{\frac{1}{2}}p^{\frac{1}{2}}).$$

This square root type deviate was studied in section 3c.

A third approximation is found when the theorem is applied to $\underline{F}^{1/3}$; WILSON & HILFERTY (1931) discovered that $(\underline{\chi}^2_{\nu_i}/\nu_i)^{1/3}$ is asymptotically normal with expectation $1 - 2/(9\nu_i)$ and variance $2/(9\nu_i)$. The resulting approximation, derived by PAULSON (1942) for the distribution of $\underline{F}$ and by CAMP (1951) for the binomial, is

$$(4.6) \qquad \Phi\left(\frac{\{9 - (k+1)^{-1}\}\ F^{1/3} - 9 + (n-k)^{-1}}{3\{F^{2/3}(k+1)^{-1} + (n-k)^{-1}\}^{\frac{1}{2}}}\right), \text{ with } F = \frac{(k+1)q}{(n-k)p}.$$

The CAMP-PAULSON approximation (4.6) is somewhat cumbersome but very accurate (RAFF, 1956). For n = 10, p = .2, $P[\underline{y} \geq 4] = .0328$ is given by (4.6) with a relative error of +1.2 per cent. (against +7.4 per cent. for the square root type (4.5) and -26.7 per cent. for the classical normal approximation). A lengthy calculation, not reproduced here, shows that the deviate in (4.6) has the expansion

$$(4.7) \qquad \xi + \sigma^{-2}\,(1-3pq)(-\,\xi^3 + 3\xi)/108 + O(\sigma^{-3}).$$

Its error is thus $O(\sigma^{-2})$ for all p and almost all ξ, but $O(\sigma^{-3})$ for $\xi = 0$ (the median) and $\xi^2 = 3$ (tails of .042). As the term of order σ^{-3}, not explicitly given in (4.7), vanishes for p = q, the error is even $O(\sigma^{-4})$ for p = q and $\xi^2 = 0$ or 3. It seems impossible to introduce simple adjustments into (4.6) which would make it asymptotically optimal at other values than $\xi^2 = 3$ or 0.

4b. BORGES APPROXIMATION

In the spirit of CURTISS (1943) and BLOM (1954), a systematic search for a suitable transformation of the fraction of successes $\underline{y}/n$ was undertaken by BORGES (1970). He proposes the approximation

$$(4.8) \qquad \Phi(\{n+1/3\}^{\frac{1}{2}}(pq)^{-1/6}\{J(\tfrac{k+2/3}{n+1/3})-J(p)\}),$$

where

$$(4.9) \qquad J(z) = \int_0^z t^{-1/3}(1-t)^{-1/3}\,dt.$$

A table of J(z) has been computed by GEBHARDT (to be published) by a numerical integration procedure. The present author computed J(z) by termwise integration of

$$(4.10) \qquad t^{-1/3}(1-t)^{-1/3} \simeq \sum_{j=0}^{j_0} \binom{-1/3}{j} (-1)^j\, t^{j-1/3}$$

for $z \leq \frac{1}{2}$. For $\frac{1}{2} < z < 1$ one may use

$$(4.11) \qquad J(z) = 2.0533902 - J(1-z),$$

where $2.0533902 = J(1) = \{\Gamma(2/3)\}^2 / \Gamma(4/3)$, the complete beta integral. For the worst case $z = \frac{1}{2}$, integration of (4.10) with $j_0 = 13$ (i.e. 14 terms) gives an answer correct to 6 decimal places, whereas $j_0 = 8$ is sufficient for 4 place accuracy.

It might be useful if the BORGES approximation (4.8) could also be applied when no table of $J(z)$ is available. We have found from empirical study that

$$J^*(z) = 1.5\ z^{2/3}(60-17z)/(60-25z) \tag{4.12}$$

is a reasonably good approximation to $J(z)$. We give some examples, each followed by the exact values $J(z)$ in brackets: $J^*(.1) = .32766$ (.32766); $J^*(.3) = .70294$ (.70317); $J^*(.5) = 1.02452$ (1.02670). The agreement is good for small z, somewhat worse for z near .5. For $z > .5$ one may use (4.11) with J replaced by J^*. If p and k/n do not differ much (as will usually be true) replacement of J by J^* will not make much difference in (4.8), as $J(z) - J^*(z)$ increases slowly with z for $0 < z < .5$.

The approximation (4.8) is very accurate, and remains accurate when J is replaced by J^*. We have derived its series expansion from the Taylor series $J(f_0) - J(p) = \sum(f_0-p)^j\, J^{(j)}(p)/j!$, where $f_0 = (k+2/3)/(n+1/3)$. For the first factor in (4.8) we have taken the slightly more general expression $\{n+d+\delta/(pq)\}^{\frac{1}{2}}$. From lengthy calculations not reproduced here one obtains

$$\begin{aligned}&\{n+d+\delta/(pq)\}^{\frac{1}{2}}(pq)^{-1/6}\{J(f_0)-J(p)\} = \xi + \\ &+ \sigma^{-2}\ \{\xi^3(1+2pq) + \xi(-6+108\delta-30pq+108dpq)\}/216 + O(\sigma^{-3}).\end{aligned} \tag{4.13}$$

The error is $O(\sigma^{-2})$ for all p and almost all ξ, but $O(\sigma^{-3})$, for $p = q$ even $O(\sigma^{-4})$, for $\xi = 0$ and $\xi^2 = (6-108\delta+30pq-108dpq)/(1+2pq)$. Thus BORGES' choice $d = 1/3$, $\delta = 0$ is asymptotically optimal at $\xi^2 = 3$ when $p = .5$, at $\xi^2 = 3.8$ when $p = .2$, etc. One may use $d = 7/27$, $\delta = 5/108$ for optimality near $\xi^2 = 1$ and $d = 11/54$, $\delta = 1/54$ for optimality near $\xi^2 = 4$, in both cases regardless of the value of p.

4c. IMPROVED BINOMIAL OR SQUARE ROOT TYPE

Let u and v be two simple deviates as described in section 3. This subsection is devoted to improvements to the approximation $\Phi(u)$, for the skew case $p \neq q$ only (see section 5a for the case $p = q$). Just as in section II.4, one can

(i) add to u a correction $R(v)$ depending on v;
(ii) add to $\Phi(u)$ an expression of the form $\phi(u)\,R(v)$;
(iii) use u with a continuity correction (or similar adjustment) depending on v.

The explicit form of $R(v)$, or of the variable continuity correction, can be deduced from the series expansions (3.3) and (3.20).

It usually simplifies the calculations to take $u = v$. We shall choose for u and/or v either the binomial type (BI)

$$\text{(4.14)} \qquad BI = (k+c-np)/\{(n+d)pq+\delta\}^{\frac{1}{2}}$$

or the square root type (SQ)

$$\text{(4.15)} \qquad SQ = 2(k+1+b)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+b)^{\frac{1}{2}}p^{\frac{1}{2}};$$

these two are more promising, as regards simple and accurate approximation, than beta or arcsin type deviates.

Method (i) leads to

$$\text{(4.16)} \qquad \Phi(u + \sigma^{-1}(q-p)(v^2-4-12b)/12) \text{ when } u = SQ,$$

$$\text{(4.17)} \qquad \Phi(u + \sigma^{-1}(q-p)(a-v^2)/6) \qquad \text{when } u = BI;$$

in both cases v may be either BI or SQ, and in the second formula the quantity a depends on the value of c used in u = BI: for $c = \frac{1}{2}$ one has $a = 1$, for $c = p$ one has $a = 4$, and for $c = q$ one has $a = -2$. FELLER (1945) uses (4.17), for u and v both BI with $c = q$, $d = 1$, $\delta = 0$, in his derivation of error bounds for $P[j \leq \underline{y} \leq k]$ under the rather restrictive conditions $(n+1)pq > 9$, $j > (n+1)p$, $k + \frac{1}{2} \leq (n+1)p + \frac{2}{3}(n+1)pq$.

For method (ii) we restrict ourselves to the case u = v, because otherwise the calculation of the approximation would become rather lengthy. This means

$$(4.18) \qquad \Phi(u) + \sigma^{-1}\phi(u)(q-p)(u^2-4-12b)/12 \text{ when } u = SQ;$$

$$(4.19) \qquad \Phi(u) + \sigma^{-1}\phi(u)(q-p)(a-v^2)/6 \qquad \text{when } u = BI;$$

again a depends on c as described immediately under (4.17). RAFF (1955, 1956) uses (4.19) with the classical choice $c = \frac{1}{2}$, $d = \delta = 0$ in the BI-deviate u.

The third method coincides with the first when u = BI, because the continuity correction c is an additive correction. For u of square root type it means

$$(4.20) \qquad \Phi(2(k+1+B)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+B)^{\frac{1}{2}}p^{\frac{1}{2}}), \qquad \text{where } B = (v^2-4)/12,$$

and v is either BI or SQ. In section 5b we shall consider a modification of (4.20) where different corrections B are used in both terms.

Let SQ denote the square root deviate (4.15) with b = 0. FREEMAN & TUKEY (1950) propose the normal deviate $u + \sigma^{-1}(u+2p-1)M^{\frac{1}{2}}/12$, where $M = \max(p,q)$, and u = SQ for rapid significance testing (say $1.5 \leq |\xi| \leq 2.5$) or $u = SQ + \{(SQ)^2-4\}\{(np+1)^{-\frac{1}{2}}-(nq+1)^{-\frac{1}{2}}\}/12$ for good accuracy at all levels. Neither asymptotic expansion nor numerical investigation has given us any reason to support their view that the stated approximations should be used.

4d. COMPARISON OF BETTER APPROXIMATIONS

From the expansions (3.3) for BI and (3.20) for SQ one can derive the asymptotic expansions of the approximations described in the preceding subsection. The leading term of the error is always of order σ^{-2}; for a uniform notation we shall introduce the values of the coefficients ζ_5, ζ_3 and ζ_1

such that the error is

$$(4.21) \qquad \sigma^{-2}\phi(\xi)(\zeta_5\xi^5 + \zeta_3\xi^3 + \zeta_1\xi)/72 + O(\sigma^{-3}).$$

For (4.18), one has $\zeta_5 = \frac{1}{4}-pq$, $\zeta_3 = -1-6b+7pq+24bpq$, $\zeta_1 = -6+36b^2-6pq-72bpq+$ $-144b^2pq$; for (4.19), $\zeta_5 = 1-4pq$, $\zeta_3 = -5-2a+14pq+8apq$ and $\zeta_1 = -2+4a+a^2+$ $-36\delta+14pq-16apq-4a^2pq-36dpq$, where a is determined by c as described below (4.17); b, c, d and δ are the adjustments occurring in u = SQ and u = BI respectively. One finds that (4.18) will generally have a smaller leading term of the error than (4.19). For all other approximations of subsection 4c, one has $\zeta_5 = 0$. We list the values of ζ_3 and ζ_1 in Table 4.1, where we include also the similar values for the BORGES and CAMP-PAULSON approximations, cf. (4.13) and (4.7).

The coefficients ζ_3 and ζ_1 in Table 4.1 vary with pq, and some of them with b, d and δ. This makes it somewhat difficult to decide which line of the Table contains low coefficients (and thus gives a low absolute value of the leading term of the error). Nevertheless one can see that the CAMP-PAULSON and BORGES approximations will usually give smaller errors than the improved simple ones. Among the latter, (4.16) with v = SQ and (4.20) with v = BI have a fairly low coefficient ζ_3 for all p. Comparing the former, one finds that ζ_3 is larger for BORGES when p is near $\frac{1}{2}$, for CAMP-PAULSON when p is near zero (or one).

The amount of computation involved, and the accuracy found from our numerical investigation must also be taken into account (although they are both difficult to measure). We then arrive at the conclusion that CAMP-PAULSON is slightly better than BORGES for all p when n is small (say 20 or less) and for roughly $.2 < p < .8$ when $n \geq 50$ (say). They are both generally better, but also more cumbersome, than (4.16) with v = SQ, $b = -\frac{1}{4}$, i.e.

$$(4.22) \qquad \Phi(u + \sigma^{-1}(q-p)(u^2-1)/12),$$
$$\text{where } \sigma = (npq)^{\frac{1}{2}} \text{ and } u = 2(k+\tfrac{3}{4})^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k-\tfrac{1}{4})^{\frac{1}{2}}p^{\frac{1}{2}}.$$

The remaining approximations listed in Table 4.1 are usually less accurate.

TABLE 4.1

Values of ζ_3 and ζ_1 such that the error of the approximations is given by (4.21) with $\zeta_5 = 0$. For (4.18) and (4.19) see the text below (4.21).

	approximation	ζ_3	ζ_1
(4.6)	CAMP-PAULSON	-2/3+2pq	2-6pq
(4.13)	BORGES	1/3+2pq/3	-2+36δ-10pq+36dpq
(4.16)	v = SQ	1-pq	-10-24b+10pq+24bpq
(4.16)	v = BI, c = ½	4-13pq	-16-36b+34pq+72bpq
(4.16)	v = BI, c = p	4-13pq	-22-36b+58pq+72bpq
(4.16)	v = BI, c = q	4-13pq	-10-36b+10pq+72bpq
(4.17)	v = SQ	1-10pq	-10-36δ-24b+46pq-36dpq+96bpq
(4.17)	v = BI, c = ½, a = 1	-5+14pq	2-36δ-2pq-36dpq
(4.17)	v = BI, c = p, a = 4	-5+14pq	14-36δ-50pq-36dpq
(4.17)	v = BI, c = q, a = -2	-5+14pq	-10-36δ+46pq-36dpq
(4.20)	v = SQ	-2+5pq	2+12b-14pq-48bpq
(4.20)	v = BI, c = ½	1-7pq	-4+10pq
(4.20)	v = BI, c = p	1-7pq	-10+34pq
(4.20)	v = BI, c = q	1-7pq	2-14pq

This holds also for (4.18), and even more for (4.19). Of course there are exceptions for certain values of p and/or ξ, but generally speaking it seems sensible to use (4.22) when a relatively simple and CAMP-PAULSON (4.6) when a more accurate approximation is sought from the class with error $O(\sigma^{-2})$. The BORGES approximation (4.8) might be used instead of CAMP-PAULSON, especially for large n and p far from ½, with use of J^* or of special tables for $(pq)^{-1/6}$ and J. The special cases considered by RAFF (4.19, u = BI, c = ½, d = δ = 0) and FELLER (4.17, u = v = BI, c = q, d = 1, δ = 0) have decidedly larger errors, and are not essentially easier.

Linear combination of two simple approximations can also lead to an error $O(\sigma^{-1})$. In our investigation, the most promising combination turned out to be $\Phi(\{(k+\frac{1}{2}-np)\sigma^{-1} + 2u\}/3)$, with u given by (4.22). It is usually no better than (4.22) itself, and worse than CAMP-PAULSON (4.6).

5. VERY ACCURATE NORMAL APPROXIMATIONS

This sections contains a brief discussion of approximations for which the error is of order σ^{-3} or even smaller. For $p \neq q$ this can only be obtained at the cost of lengthy calculations: we discuss in subsection 5b approximations suggested by PEIZER & PRATT and BOLSHEV, and some improved simple deviates. But first we shall consider the case $p = q$, where an error of order σ^{-4} can be reached without much effort.

5a. SYMMETRIC CASE $p = q = \frac{1}{2}$

Binomial distributions with $p = q = \frac{1}{2}$ are frequently used, e.g. in applications of the sign test. For this reason it may be worthwhile to point out some very accurate and relatively simple approximations for this case.

For $p = q = \frac{1}{2}$, $c = \frac{1}{2}$ and $d = \delta = 0$, the simple binomial deviate (3.3) becomes

$$\text{(5.1)} \qquad BI = (2k+1-n)n^{-\frac{1}{2}}.$$

The simple square root type deviate (3.20) becomes for $p = q = \frac{1}{2}$

$$\text{(5.2)} \qquad SQ = (2k+2+2b)^{\frac{1}{2}} - (2n-2k+2b)^{\frac{1}{2}}.$$

From the series expansions (3.4) and (3.22) it follows that the following approximations to $G_{n,\frac{1}{2}}(k)$ all have an error of order σ^{-4}:

$$\text{(5.3)} \qquad \Phi(u + (v^3-v)/(12n)), \qquad u = BI,\ v = BI \text{ or } SQ;$$

$$\text{(5.4)} \qquad \Phi(u - (v^3-10v-24bv)/(24n)), \quad u = SQ,\ v = BI \text{ or } SQ;$$

$$\text{(5.5)} \qquad \Phi((2k+2+\beta)^{\frac{1}{2}} - (2n-2k+\beta)^{\frac{1}{2}}), \quad \beta = \{(2k+1-n)^2-10n\}/(12n);$$

$$\text{(5.6)} \qquad \Phi((2k+2+\beta)^{\frac{1}{2}} - (2n-2k+\beta)^{\frac{1}{2}}), \quad \beta = \{(SQ)^2-10\}/12.$$

Numerical investigation shows that (5.5) is very accurate, and its calculation is reasonably simple. Slightly better, but also more cumbersome, is (5.4) with u = v = SQ and $b = -\frac{1}{4}$, i.e.

$$(5.7) \qquad \Phi(u - (u^3-4u)/(24n)), \text{ where } u = (2k+3/2)^{\frac{1}{2}} - (2n-2k-\tfrac{1}{2})^{\frac{1}{2}}.$$

In the example n = 10, $p = \frac{1}{2}$, the approximation (5.5) gives $P[\underline{y} \leq 1] = .0107$ with a relative error of -.6 per cent., and $P[\underline{y} \leq 2] = .0547$ with a relative error of -.3 per cent. For n = 20, $p = \frac{1}{2}$, the relative tail error of (5.5) lies between 0 and -.1 per cent. for all probabilities between .001 and .999. As explicit calculation of

$$(5.8) \qquad G_{n,\frac{1}{2}}(k) = \sum_{j=0}^{k} \binom{n}{j} 2^{-n},$$

or its complement, is easily carried out for n < 10, we may say that any symmetric binomial distribution can be handled without access to binomial tables: use (5.8) or its complement for n < 10, (5.5) for $10 \leq n \leq 100$, and $\Phi((2k+7/4)^{\frac{1}{2}} - (2n-2k-\frac{1}{4})^{\frac{1}{2}})$ for n > 100. Then the relative tail error will lie between -1 per cent. and +1 per cent. for all probabilities between .001 and .999.

5b. SKEW CASE $p \neq q$

For $p \neq \frac{1}{2}$ there are no easy approximations with error of order σ^{-3} or lower. The formulae in this subsection will usually be too cumbersome for hand calculation. They might be programmed for use in an electronic computer, for situations where n is too large for direct summation of the binomial terms.

PEIZER & PRATT (1968) propose a general normal approximation to beta and related distributions. In the binomial situation, its deviate can be written as

$$(5.9) \qquad \{k + 2/3 - (n+1/3)p\} \times \\ \times\{1 + qT(\frac{k+\frac{1}{2}}{np}) + pT(\frac{n-k-\frac{1}{2}}{nq})\}^{\frac{1}{2}} \{(n+1/6)pq\}^{-\frac{1}{2}},$$

where $T(z) = (1 - z^2 + 2z \log z)(1 - z)^{-2}$, with $T(1) = 0$ by continuity, is a function for which a brief table is given in the reference just mentioned. There one also finds the series expansion of (5.9), which is

$$(5.10) \qquad \xi + \sigma^{-3}\ (q-p)\{\xi^2(1+23pq) + 32+16pq\}/1620 + O(\sigma^{-4}).$$

The absence of terms of order σ^{-1} and σ^{-2}, and even of a term proportional to $\sigma^{-3}\xi^4$, explains the extraordinary accuracy of this approximation. PEIZER & PRATT propose a further refinement, viz. the addition of

$$(5.11) \qquad .02\left(\frac{q}{k+1} - \frac{p}{n-k} + \frac{q-\frac{1}{2}}{n+1}\right)$$

to the first factor between curly brackets in (5.9). We shall consider a modification of (5.9) which includes this refinement, viz.

$$(5.12) \qquad \begin{aligned} &\{k + \tfrac{2}{3} + \varepsilon_1\{\frac{q}{k+1} - \frac{p}{n-k}\} + \varepsilon_2\,\frac{q-\frac{1}{2}}{n+1} - (n+\tfrac{1}{3})p\}\times \\ &\times\{1 + qT(\frac{k+\frac{1}{2}}{np}) + pT(\frac{n-k-\frac{1}{2}}{nq})\}^{\frac{1}{2}}\ \{(n+1/6)pq\}^{-\frac{1}{2}}. \end{aligned}$$

The error of (5.12) is

$$(5.13) \qquad \sigma^{-3}\ (q-p)\{\xi^2+32-1620\varepsilon_1 + pq(23\xi^2+16-810\varepsilon_2)\}/1620 + O(\sigma^{-4}).$$

For asymptotic optimality at $\xi = 0$ (the median), one should choose $\varepsilon_1 = \varepsilon_2 = 8/405 \simeq .02$ (as proposed by PEIZER & PRATT). Asymptotic optimality at $\xi^2 = 1$ (tails of .16) means $\varepsilon_1 = 11/540 \simeq .02$, $\varepsilon_2 = 13/270 \simeq .05$, and at $\xi^2 = 4$ (tails of .023) one must take $\varepsilon_1 = 1/45 \simeq .02$ and $\varepsilon_2 = 2/15 \simeq .13$. We recall that this holds for $p \neq q$; for $p = q = \frac{1}{2}$ the error is $O(\sigma^{-4})$, but then (5.5) and (5.7) are easier and more accurate.

Numerical investigations (see the Tables in section 7) confirm the remarkable accuracy of (5.9), and even more of (5.12) with the just mentioned choices of ε_1 and ε_2. In some examples with $n = 30$, $p < .3$, the error was roughly halved by taking $\varepsilon_1 = .02$, $\varepsilon_2 = .13$ instead of $\varepsilon_1 = \varepsilon_2 = .02$. However the error is already very small, even for (5.9) without refinements.

Exceptions are cases with very small k. Especially for k = 0, the approximation is a lot worse, regardless of the values of ε_1 and ε_2 that are used. This is not very serious: one may calculate the exact value q^n when n is not too large, and for large n the value k = 0 is highly improbable unless p is so small that even a simple Poisson approximation can be used (see section 6).

We refer to PEIZER & PRATT (1968) for more information about the somewhat cumbersome, but extremely accurate normal approximation (5.9) and its refinement by means of (5.11).

BOLSHEV a.o. (1961) give a very accurate normal approximation, based on the expansion by WISHART (1957) of the distribution function of FISHER's z. It has the form

$$\Phi(u) + \phi_1(u,v) + (n+1)(k+1)^{-1}(n-k)^{-1}\phi_2(u,v), \tag{5.14}$$

where

$$\begin{aligned} u &= (k+1)^{\frac{1}{2}}(n-k)^{\frac{1}{2}}(n+1)^{-\frac{1}{2}}\log\left\{\frac{(k+1)q}{(n-k)p}\right\}, \\ v &= (2k+1-n)(k+1)^{-\frac{1}{2}}(n-k)^{-\frac{1}{2}}(n+1)^{-\frac{1}{2}}, \end{aligned} \tag{5.15}$$

and ϕ_1 and ϕ_2 are linear combinations of derivatives of $\Phi(u)$, multiplied by powers of v. A two page table of ϕ_1 and ϕ_2 is given by BOLSHEV a.o. (1961). The absolute error of (5.14) never exceeds 5.10^{-5} provided that $\min(k+1,n-k) > 20$ and $\max(k+1,n-k) > 50$. Numerical experience shows that $\Phi(u)$ itself is not very accurate. This follows also from the series expansion

$$\begin{aligned} u &= \xi + \sigma^{-1}(q-p)(\xi^2 + 2)/6 + \\ &+ \sigma^{-2}\{\xi^3(-4-2pq) + \xi(-2+14pq)\}/72 + O(\sigma^{-3}). \end{aligned} \tag{5.16}$$

The complete formula (5.14) could be of some interest for computer programming, but is wholly inadequate for hand calculation.

Approximations with an error of order σ^{-3} may also be found from

$$\Phi(2\{(k+1)q + A\}^{\frac{1}{2}} - 2\{(n-k)p + B\}^{\frac{1}{2}}), \tag{5.17}$$

if the adjustments A and B are allowed to depend on some simple deviate v. A slight modification of (3.19) shows that

$$\begin{aligned} &2\{(k+1)q + A\}^{\frac{1}{2}} - 2\{(n-k)p + B\}^{\frac{1}{2}} = \xi + \\ &+ \sigma^{-1} \{(q-p)(4 - \xi^2) + 12(A-B)\}/12 + \\ &+ \sigma^{-2} \{\xi^3(2-5pq) + \xi(-14+26pq-36Aq-36Bp)\}/72 + \\ &+ 0(\sigma^{-3}). \end{aligned} \tag{5.18}$$

It follows after some calculation that the terms of order σ^{-1} and σ^{-2} vanish if one uses

$$\begin{aligned} &A = (1+2p-5p^2)v^2/36 + (-5-p+7p^2)/18 + (1-4pq)b/3, \\ &B = (-2+8p-5p^2)v^2/36 + (1-13p+7p^2)/18 + (1-4pq)b/3, \\ &\text{where } v = 2(k+1+b)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+b)^{\frac{1}{2}}p^{\frac{1}{2}}, \end{aligned} \tag{5.19}$$

or

$$\begin{aligned} &A = (4-10p+7p^2)v^2/36 + (-11+17p-5p^2)/18 + (q-p)c/3, \\ &B = (1-4p+7p^2)v^2/36 + (-5+5p-5p^2)/18 + (q-p)c/3, \\ &\text{where } v = (k+c-np)/\sigma. \end{aligned} \tag{5.20}$$

It follows from Table 4.1 that an improved version of (4.16) with u = v = = SQ, viz.

$$\begin{aligned} &\Phi(u + \sigma^{-1}(q-p)(u^2-4-12b)/12 - \sigma^{-2}(1-pq)(u^3-10u-24bu)/72), \\ &\text{where } u = 2(k+1+b)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k+b)^{\frac{1}{2}}p^{\frac{1}{2}}, \end{aligned} \tag{5.21}$$

also has an error of order σ^{-3}. In a numerical investigation, various values for b and c were tried in (5.21) and in (5.17) combined with (5.19) or (5.20). They are substantially more accurate even than (2.1) with the term of order σ^{-3} included; this confirms the experience of others that the CORNISH - FISHER formulae do not provide much improvement in accuracy when more than two terms are added. It is difficult to decide which one of the approximations (5.17) and (5.21) is best, and with which choice of A, B, b or c. As a whole one could say that (5.17) with (5.20) and $c = \frac{1}{2}$, i.e.

$$\Phi\ (2\{(k+1)q + A\}^{\frac{1}{2}} - 2\{(n-k)p + B\}^{\frac{1}{2}}),$$

$$(5.22) \qquad A = \frac{(4-10p+7p^2)(k+\frac{1}{2}-np)^2}{36npq} - \frac{8-11p+5p^2}{18},$$

$$B = \frac{(1-4p+7p^2)(k+\frac{1}{2}-np)^2}{36npq} - \frac{2+p+5p^2}{18},$$

is relatively simple and relatively accurate. In accuracy it comes in many cases between the PEIZER-PRATT formula (5.9) and its refinement with (5.11). Without machines or tables its calculation may be easier.

In section 7 one finds numerical values of the errors of the normal (sections 3-5) and Poisson (section 6) approximations to the binomial distribution function and a brief advice summarizing the results.

6. POISSON APPROXIMATIONS

A previous publication (MOLENAAR, 1969) contains a detailed investigation of Poisson approximations to the binomial distribution function $G_{n,p}(k)$. This section summarizes its results. Unless the contrary is stated, they are valid for $p < .5$ and also for $p = .5$ and $k \leq \frac{1}{2}n$. In the remaining cases one should replace p by 1-p and k by n-k-1 in all formulae, and subtract the result from 1. This operation will be called *reversal*. As defined in I (1.1), $F_\lambda(k)$ denotes the Poisson distribution function. For the asymptotic results in this section, we shall assume the existence of positive constants μ_0 and K_0 such that $n \to \infty$, $p \to 0$, $np \to \mu_0$, $0 \leq k \leq K_0$.

It is common knowledge that the Poisson distribution with expectation np is an approximation to a binomial distribution with small p. The value of n has hardly any influence on the accuracy. However, both tails of the binomial distribution are rather seriously overestimated unless p is really very small, as is illustrated in Table 6.1. The examples illustrate the conservatism of the approximation: in hypothesis testing with e.g. α = .025 or .05 some significant results are not recognized as such, but no non-significant result is ever called significant (see ANDERSON & SAMUELS, 1967).

TABLE 6.1

Accuracy of the Poisson approximation with expectation np to the binomial distribution function or its complement.

n	p	Binomial tail	Poisson (np) approx.
5	.05	$P(\underline{y} \geq 2)$=.023	.026
30	.1	$P(\underline{y} = 0)$=.042	.050
40	.1	$P(\underline{y} \geq 8)$=.042	.051
40	.2	$P(\underline{y} \geq 13)$=.043	.064
70	.1	$P(\underline{y} \geq 12)$=.044	.053
100	.1	$P(\underline{y} \leq 4)$=.024	.029
100	.2	$P(\underline{y} \leq 13)$=.047	.066
125	.05	$P(\underline{y} \geq 2)$=.048	.052

The accuracy can be improved by the addition of correction terms, but still better results are obtained by replacing np by a parameter depending not only on n and p but also on k. Table 6.2 lists some parameters λ_i for which $F_{\lambda_i}(k)$ is close to $G_{n,p}(k)$. Generally speaking both accuracy and computational effort increase from the top of the Table to the bottom. RAFF (1955, 1956) mentions the Gram-Charlier type approximation

$$F_{np}(k) + \tfrac{1}{2}p(k-np)e^{-np}(np)^k/k! \tag{6.1}$$

This is roughly as accurate as F_{λ_2} or F_{λ_4}, and not easier to compute, even

TABLE 6.2

Parameters λ_i such that $G_{n,p}(k) \simeq F_{\lambda_i}(k)$, i.e. the probability of k or less events in a Poisson distribution with expectation λ_i, is an approximation to the same probability in the binomial (n,p) distribution.

$$\lambda_1 = np$$
$$\lambda_2 = (2n-k)p/(2-p)$$
$$\lambda_3 = \{(2-\tfrac{1}{2}p)n-k\}p/(2-3p/2)$$
$$\lambda_4 = -(n-\tfrac{1}{2}k)\log q$$
$$\lambda_5 = \{(12-2p)n-7k\}np/\{(12-8p)n-k+k/n\}$$
$$\lambda_6 = -(24n^2-24nk+5k^2)(\log q)/(24n-12k)$$
$$\lambda_7 = \lambda_2\left[1+(2\lambda_2^2-k\lambda_2-k^2-2k)/\{6(2n-k)^2\}\right]$$
$$\lambda_8 = \lambda_4\left[1-k(k+2+\lambda_4)/\{6(2n-k)^2\}\right]$$
$$\lambda_9 = \lambda_2\left[1-(2\lambda_2^2-k\lambda_2-k^2-2k)/\{6(2n-k)^2\}\right]^{-1}$$
$$\lambda_{10} = \lambda_4\left[1+k(k+2+\lambda_4)/\{6(2n-k)^2\}\right]^{-1}$$

when a table of individual terms of the Poisson distribution is available. A similar refinement to F_{λ_4}, and F_{λ_4} itself, were first proposed by WISE (1946, 1950), who gives also an inverse result from which F_{λ_8} and $F_{\lambda_{10}}$ can be deduced. F_{λ_2}, F_{λ_9} and a refinement of F_{λ_2} similar to (6.1), were derived in a series of papers by BOLSHEV (1961, 1963, 1964). The present author proposed (MOLENAAR, 1969): λ_3 (a refinement to λ_2), λ_6 (a refinement to λ_4) and λ_5. The latter has an error of a smaller order of magnitude than λ_1, λ_2, λ_3, λ_4 or λ_6, but of a higher order of magnitude than λ_7 up to λ_{10}.

In our opinion, λ_2 is excellent for simple work, and λ_5 can be used when greater accuracy is desired. In both cases reversal (interchange of successes and failures) should be applied for $p > \frac{1}{2}$ and for $p = \frac{1}{2}$ and $k > \frac{1}{2}n$, as indicated at the beginning of this section. Although it is difficult to formulate a simple rule, reversal might also be successful when $\frac{1}{2}(1-n^{-\frac{1}{2}}) < p < \frac{1}{2}$, at least for small right hand tails. The use of λ_7 up to λ_{10}, or of F_{λ_2} or F_{λ_4} with a correction containing an individual Poisson term, leads to very accurate results, but asks for a lot of computation.

There is not much difference between them: λ_9 is a little better for p near $\frac{1}{2}$, and λ_8 a little better for p near 0. The approximations F_{λ_1}, F_{λ_2} and F_{λ_5} are conservative, i.e. they overestimate binomial tails.

The Poisson parameter λ_3 is slightly easier than λ_5. When F_{λ_3} is used instead of F_{λ_5}, one finds for some values of p that it is rather more accurate for probabilities near 0 than for probabilities near 1. In such situations it is useful to apply reversal in other cases than described above. It was empirically found that reversal makes λ_3 better:

(a) for $k > \frac{1}{2}n$ when $.2+.016n^{\frac{1}{2}} < p < .8-.016n^{\frac{1}{2}}$ (this means e.g. $.25 < p < .75$ for $n = 20$, and $.4 < p < .6$ for $n = 160$).

(b) for all k when $p \geq .8-.016n^{\frac{1}{2}}$.

With this modified reversal rule, F_{λ_3} becomes about as accurate as F_{λ_5}.

Furthermore, one can show that $F_{\lambda_4}(k)$ is more accurate than $F_{\lambda_3}(k)$ if and only if $k > np/3$. Unless np is rather small, this situation occurs only for very small probabilities. Reversal makes F_{λ_4} better for about the same cases (a) and (b) as above. Consequently the reversed form of $F_{\lambda_4}(k)$ should be used instead of the reversed form of $F_{\lambda_3}(k)$ when before reversal (a) or (b) holds and $k > nq/3$; unless nq is small this occurs only for probabilities very close to 1.

We now shall give a brief description of the asymptotic investigations contained in MOLENAAR (1969). Monotonicity arguments guarantee the existence and uniqueness of the *exact Poisson parameter* λ_0 defined by $F_{\lambda_0}(k) = G_{n,p}(k)$. Its explicit solution from this transcendental equation is impossible, but by passing to the corresponding incomplete gamma and beta functions, cf. II (1.4) and III (1.2), an expansion for λ_0 in powers of n^{-1} can be derived for $n \to \infty$, np and k bounded. The first terms are

$$(6.2) \qquad \lambda_0 = np[1 + n^{-1}(np - k)/2 + n^{-2}(8n^2p^2-7npk-k^2-2k)/24 + O(n^{-3})].$$

In CAMPBELL (1923) np is similarly expressed as a function of λ_0. WISE (1950) also expresses np in λ_0, but expands in powers of $(2n-k)^{-1}$, which simplifies the formulae and improves convergence.

From a comparison of (6.2) to similar expansions for λ_2 up to λ_6, the following asymptotic conclusions, (formulated only for $p \leq .5$, without considering reversal) are derived in MOLENAAR (1969). We shall use "iff" for "if and only if".

(i) λ_5 is superior to any other λ_i, $i \leq 6$, except for $k = 0$ (where λ_4 and λ_6 do trivially give the exact probability q^n).

(ii) λ_4 is superior to λ_3 for $k < np/3$, and λ_4 is never superior to λ_6.

(iii) λ_3 is superior to λ_6 in a middle region $.38np < k < 2.62np$ containing most of the distribution.

(iv) λ_3 is superior to λ_2, except in a narrow strip $-1 + np/4 + (9n^2p^2/16 - np/2 + 1)^{\frac{1}{2}} < k < np$.

(v) λ_4 is superior to λ_2 iff $k < -1 - \frac{1}{2}np + (5n^2p^2/4 + np + 1)^{\frac{1}{2}}$, i.e. in the left hand tail.
λ_2 is superior to λ_6 iff $-np + 2 + (3n^2p^2 + 4np + 4)^{\frac{1}{2}} < k < 2^{\frac{1}{2}}np$, which is rather a wide region.

(vi) $F_{\lambda_1}(k)$ overestimates $G_{n,p}(k)$ iff $k < np$; $F_{\lambda_2}(k)$ overestimates $G_{n,p}(k)$ iff $k < -1 - \frac{1}{2}np + (9n^2p^2/4 + np + 1)^{\frac{1}{2}}$; $F_{\lambda_5}(k)$ overestimates $G_{n,p}(k)$ for $k \leq \max(np-1, 0)$ and underestimates it for $k \geq np$.

The paper just mentioned gives a graph for $0 \leq np \leq 9$ of the functions of np used in these asymptotic conclusions. It reports on the numerical verification of the conclusions for 84 parameter pairs (n,p) with $5 \leq n \leq 300$ and $.01 \leq p \leq .50$; exceptions turn out to be rare. The paper also contains extensive tables of relative tail errors.

7. GENERAL ADVICE AND NUMERICAL INFORMATION

This section opens with a general advice on normal and Poisson approximations to the binomial distribution function (Table 7.1). A simple recommendation inevitably contains some subjective elements; we refer to the summary and to Chapter I for a general outline of our view. Next the relative tail error I (2.3) is tabled for $p = .5$, $.4$, $.2$ and $.05$ and various n, for some approximations selected from the previous sections. Finally we give two graphs illustrating the maximal absolute error of the best simple

(text continued on page 115)

TABLE 7.1. Advice for approximation to the binomial distribution function $G_{n,p}(k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^j$. In most statistical applications, accurate approximation to probabilities between .005 and .05 or between .95 and .995 will be essential. In such cases, one may use the suggestions marked "for tails". Φ denotes the standard normal distribution function I (2.1).

For quick work, use

$$\Phi(2(k+1)^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k)^{\frac{1}{2}}p^{\frac{1}{2}}) \text{ for tails,}$$

$$\Phi((4k+3)^{\frac{1}{2}}q^{\frac{1}{2}} - (4n-4k-1)^{\frac{1}{2}}p^{\frac{1}{2}}) \text{ for values between .05 and .93.}$$

When p is close to $\frac{1}{2}$, say $.25 \leq p \leq .75$ for $n = 3$ or $.40 \leq p \leq .60$ for $n = 30$ or $.46 \leq p \leq .54$ for $n = 300$, it is still better to use

$$\Phi((4k+3)^{\frac{1}{2}}q^{\frac{1}{2}} - (4n-4k-1)^{\frac{1}{2}}p^{\frac{1}{2}}) \text{ for tails,}$$

$$\Phi((4k+2\tfrac{1}{2})^{\frac{1}{2}}q^{\frac{1}{2}} - (4n-4k-1\tfrac{1}{2})^{\frac{1}{2}}p^{\frac{1}{2}}) \text{ for values between .05 and .93.}$$

When cumulative Poisson tables are available and p is small, say $p \leq .4$ for $n = 3$ or $p \leq .3$ for $n = 30$ or $p \leq .2$ for $n = 300$, use *)

$$\sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! \text{ with } \lambda = (2n-k)p/(2-p).$$

For approximations accurate near probabilities of α and $1-\alpha$, but somewhat rough elsewhere, consult section 3d (p.88).

For accurate approximation, use (5.22) given on page 105 for $p \neq \frac{1}{2}$, and (5.5) given on page 100 for $p = \frac{1}{2}$. When cumulative Poisson tables are available and p is small, say $p \leq .4$ for $n = 3$ or $p \leq .24$ for $n = 30$ or $p \leq .12$ for $n = 300$, use *)

$$\sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! \text{ with } \lambda = \frac{(12-2p)n-7k}{(12-8p)n-k+k/n}\,np.$$

*) In cases where p is only just "small" enough in the sense described above, it may be useful to know that the Poisson approximations tend to be somewhat less accurate for right hand tails (distribution function near 1). When (1-p) is "small", the Poisson approximations can be used after interchanging successes and failures.

TABLE 7.2. Event $\underline{y} \le k$, exact binomial probability and relative tail error in per cent. for some normal and Poisson approximations. Case $p = .5$. The relative tail error for $\underline{y} \ge n-k-1$ equals the relative tail error for $\underline{y} \le k$, provided that for Poisson approximations successes and failures are interchanged. Example: for $G_{10,.5}(2) = P[\underline{y} \le 2] = .0547$, approximation (3.20) with $b = 0$ has a relative error of +10.65 per cent., it gives $(1+.1065)\times.0547 = .0605$.

Event	Probability	Peizer-Pratt=(5.12) $\varepsilon_1=0$ $\varepsilon_2=0$	Peizer-Pratt $\varepsilon_1=.02$ $\varepsilon_2=.02$	Peizer-Pratt $\varepsilon_1=.02$ $\varepsilon_2=.13$	corrected squareroot (5.5)	corrected squareroot (5.7)	Camp-Paulson (4.6)	simple square root (3.20) b=-3/8	simple square root (3.20) b=-1/4	simple square root (3.20) b=0	binom. u_b (2.2)	Poisson approximations for λ_i Table 6.2, p.107: $\lambda_1=np$	λ_2	λ_5	λ_3
n = 10	p = .5														
$\underline{y} \le 1$	.0107	-1.88	-1.21	-1.21	-.56	-.03	+.55	-12.74	-3.41	+15.99	+25.01	+276.34	+21.25	+3.99	-3.76
$\underline{y} \le 2$	.0547	-.64	-.37	-.37	-.34	-.10	-.22	-2.57	+1.88	+10.65	+4.09	+127.94	+13.31	+2.55	-1.96
$\underline{y} \le 3$	.1719	-.21	-.11	-.11	-.16	-.15	-.18	-.10	+1.66	+5.05	-.28	+54.20	+6.77	+1.39	-1.11
$\underline{y} \le 4$	.3770	-.04	-.02	-.02	-.04	-.05	-.05	+.08	+.46	+1.18	-.28	+16.86	+1.89	+.46	-.97
n = 20	p = .5														
$\underline{y} \le 3$	.0013	-.86	-.57	-.57	-.07	-.49	+2.96	-17.96	-10.99	+3.90	+41.66	+702.23	+37.38	+6.22	-4.59
$\underline{y} \le 4$	.0059	-.48	-.30	-.30	-.10	-.02	+.93	-8.53	-3.56	+6.71	+17.67	+395.06	+28.62	+4.80	-2.90
$\underline{y} \le 5$	.0207	-.26	-.15	-.15	-.09	+.02	+.15	-3.47	-.15	+6.55	+6.72	+224.17	+20.93	+3.57	-1.71
$\underline{y} \le 6$	.0577	-.14	-.07	-.07	-.07	-.02	-.10	-1.07	+.99	+5.08	+1.91	+125.71	+14.32	+2.51	-.98
$\underline{y} \le 7$	.1316	-.07	-.03	-.03	-.05	-.04	-.12	-.15	+1.00	+3.26	+.14	+67.36	+8.82	+1.61	-.64
$\underline{y} \le 8$	.2517	-.03	-.01	-.01	-.03	-.03	-.08	+.07	+.60	+1.62	-.22	+32.22	+4.45	+.86	-.59
$\underline{y} \le 9$	.4119	-.01	-.00	-.00	-.01	-.01	-.02	+.03	+.16	+.41	-.09	+11.17	+1.23	+.27	-.71
n = 50	p = .5														
$\underline{y} \le 14$	.0013	-.12	-.08	-.08	-.02	-.07	+1.14	-6.84	-4.26	+1.03	+14.49	+853.21	+47.01	+6.89	-1.36
$\underline{y} \le 16$	.0077	-.06	-.04	-.04	-.01	+.00	+.33	-2.78	-1.05	+2.46	+5.62	+391.93	+32.18	+4.94	-.30
$\underline{y} \le 18$	.0325	-.03	-.02	-.02	-.01	+.00	+.02	-.85	+.21	+2.33	+1.67	+183.60	+20.41	+3.29	+.15
$\underline{y} \le 20$	.1013	-.01	-.01	-.01	-.01	-.01	-.05	-.13	+.44	+1.56	+.22	+83.08	+11.39	+1.93	+.15
$\underline{y} \le 22$	.2399	-.00	-.00	-.00	-.00	-.00	-.04	+.03	+.26	+.71	-.08	+32.34	+4.91	+.87	-.12
$\underline{y} \le 24$	.4439	-.00	-.00	-.00	-.00	-.00	-.01	+.01	+.04	+.10	-.02	+6.65	+.73	+.14	-.45
n = 100	p = .5														
$\underline{y} \le 36$	.0033	-.02	-.02	-.02	-.01	-.01	+.32	-2.21	-1.16	+.95	+4.47	+615.94	+43.64	+6.33	+1.27
$\underline{y} \le 39$	.0176	-.01	-.00	-.00	-.00	+.00	+.06	-.75	-.09	+1.23	+1.50	+266.88	+27.45	+4.23	+1.25
$\underline{y} \le 42$	.0666	-.00	-.00	-.00	-.00	-.00	-.02	-.16	+.21	+.94	+.30	+115.45	+15.46	+2.52	+.88
$\underline{y} \le 45$	.1841	-.00	-.00	-.00	-.00	-.00	-.03	+.01	+.17	+.50	-.02	+44.96	+7.04	+1.21	+.33
$\underline{y} \le 48$	.3822	-.00	-.00	-.00	-.00	-.00	-.01	+.01	+.05	+.12	-.02	+11.17	+1.68	+.30	-.18

TABLE 7.3. Event $\underline{y} \leq k$ or $\underline{y} \geq k+1$, exact binomial probability and relative tail error in per cent. for some normal and Poisson approximations. Case p = .4

Event	Proba-bility	Peizer-Pratt=(5.12) $\varepsilon_1=0$ $\varepsilon_2=0$	$\varepsilon_1=.02$ $\varepsilon_2=.02$	$\varepsilon_1=.02$ $\varepsilon_2=.13$	corrected square roots (5.22)	(4.22)	Camp-Paulson (4.6)	square root (3.20) b=-1/4	(3.20) b=0	binom. u_b (2.2)	Poisson approximations for λ_i Table 6.2, p.107 $\lambda_1=np$	λ_2	λ_5	λ_3
n = 5	p = .4													
$\underline{y} \leq 0$	.0778	-4.89	-2.97	-2.64	-2.37	+3.87	-1.10	+.96	+29.12	+9.89	+74.04	+5.56	+.87	-1.72
$\underline{y} \leq 1$	.3370	-.64	-.22	-.03	-.56	+.86	+.02	+2.25	+9.01	-3.83	+20.49	+1.66	+.36	-.83
$\underline{y} \geq 3$	.3174	+.09	-.08	-.26	+.04	+1.72	-.17	+.39	-1.96	+2.08	+1.85	+1.85	+.24	-1.55
$\underline{y} \geq 4$	.0870	-.74	-.63	-.92	-.91	+3.87	-.09	+6.05	+9.35	-1.82	+64.15	+15.82	+4.11	-2.18
$\underline{y} \geq 5$	.0102	-5.65	-4.41	-4.84	-2.10	-7.07	+.96	+10.12	+36.06	+9.76	+414.19	+81.41	+28.88	-5.34
n = 20	p = .4													
$\underline{y} \leq 2$	.0036	-1.29	-.76	-.68	-.23	-6.79	+1.87	-20.01	+.93	+66.96	+280.84	+15.29	+2.09	-3.15
$\underline{y} \leq 3$	.0160	-.66	-.35	-.29	-.13	-.76	+.16	-7.62	+7.33	+25.24	+165.52	+11.36	+1.54	-2.00
$\underline{y} \leq 4$	.0510	-.35	-.16	-.11	-.10	+1.04	-.22	-1.61	+8.52	+8.09	+95.54	+7.87	+1.08	-1.22
$\underline{y} \leq 5$	.1256	-.18	-.06	-.02	-.09	+1.09	-.19	+.71	+7.28	+1.05	+52.26	+4.89	+.69	-.76
$\underline{y} \leq 6$	.2500	-.09	-.02	+.01	-.07	+.60	-.08	+1.14	+5.20	-1.29	+25.34	+2.47	+.37	-.55
$\underline{y} \leq 7$	.4159	-.05	-.00	+.02	-.04	+.12	-.00	+.80	+3.15	-1.48	+8.91	+.66	+.11	-.49
$\underline{y} \geq 9$	.4044	+.03	-.00	-.02	+.03	+.23	-.05	-.45	-2.28	+1.32	+.75	+.75	+.10	+.75
$\underline{y} \geq 10$	.2447	+.03	-.01	-.04	+.01	+.68	-.10	+.17	-1.58	+.87	+15.82	+3.36	+.53	+1.62
$\underline{y} \geq 11$	.1275	+.01	-.01	-.05	-.03	+1.05	-.11	+1.40	+.02	-.47	+44.38	+8.03	+1.37	+.19
$\underline{y} \geq 12$	.0565	-.01	-.03	-.07	-.07	+.97	-.01	+3.24	+2.60	-2.57	+98.00	+15.80	+2.90	+.37
$\underline{y} \geq 13$	.0210	-.05	-.05	-.11	-.07	-.13	+.29	+5.55	+6.15	-4.94	+203.38	+28.39	+5.55	+.08
$\underline{y} \geq 14$	.0065	-.13	-.11	-.17	+.06	-3.04	+.96	+7.97	+10.50	-6.75	+428.63	+48.89	+10.04	-.79
$\underline{y} \geq 15$	.0016	-.27	-.21	-.28	+.43	-8.73	+2.30	+9.77	+15.18	-6.65	+970.85	+83.41	+17.69	-2.32
n = 100	p = .4													
$\underline{y} \leq 26$	.0024	-.06	-.04	-.03	-.04	-1.65	+.52	-8.84	-3.26	+22.24	+413.87	+23.14	+2.63	-.83
$\underline{y} \leq 30$	.0248	-.03	-.01	-.01	-.00	+.03	+.03	-2.25	+1.47	+5.88	+148.94	+12.62	+1.51	-.15
$\underline{y} \leq 34$	.1303	-.01	-.00	+.00	-.00	+.22	-.04	+.07	+2.26	+.34	+48.75	+5.30	+.67	-.04
$\underline{y} \leq 38$	.3822	-.00	-.00	+.00	-.00	+.05	-.01	+.36	+1.45	-.64	+8.85	+.93	+.12	-.18
$\underline{y} \geq 43$	.3033	+.00	-.00	-.00	+.00	+.10	-.02	-.19	-1.20	+.55	+11.52	+2.00	+.27	+.67
$\underline{y} \geq 47$	.0930	+.00	+.00	-.00	-.00	+.23	-.03	+.68	-.53	-.75	+63.60	+8.94	+1.26	+2.28
$\underline{y} \geq 51$	.0168	+.01	+.00	-.00	-.00	-.12	+.09	+2.85	+1.65	-4.28	+213.98	+23.03	+3.31	+5.44
$\underline{y} \geq 55$	.0017	+.00	-.00	-.01	+.04	-1.79	+.50	+6.54	+5.59	-10.04	+719.57	+49.31	+7.02	+7.77

TABLE 7.4. Event $\underline{y} \le k$ or $\underline{y} \ge k+1$, exact binomial probability and relative tail error in per cent. for some normal and Poisson approximations. Case p = .2

Event	Probability	Peizer-Pratt=(5.12) $\varepsilon_1=0$ $\varepsilon_2=0$	$\varepsilon_1=.02$ $\varepsilon_2=.02$	$\varepsilon_1=.02$ $\varepsilon_2=.13$	corrected square roots (5.22)	(4.22)	Camp-Paulson (4.6)	square root (3.20) b=-1/4	(3.20) b=0	binom. u_b (2.2)	Poisson approximations for λ_i Table 6.2, p.107 λ_1=np	λ_2	λ_5	λ_3
n = 5 p = .2														
$\underline{y} \le 0$	.3277	-3.32	-1.22	-.50	-2.75	-.09	+.22	+5.14	+27.07	-12.09	+12.27	+.46	+.03	-.19
$\underline{y} \ge 2$	.2627	+1.05	+.01	-.70	+.58	+4.27	-.53	+.07	-12.70	+9.65	+.58	+.58	+.03	+.58
$\underline{y} \ge 3$	.0579	+1.00	+.02	-1.05	-.79	+4.51	+1.18	+19.14	+4.74	-19.26	+38.64	+5.39	+.98	+3.56
$\underline{y} \ge 4$	.0067	+.64	-.16	-1.61	+6.61	-11.64	+6.44	+67.82	+54.27	-61.40	+182.56	+22.82	+6.53	+15.94
n = 20 p = .2														
$\underline{y} \le 0$	.0115	-6.14	-3.42	-3.15	-.52	-4.30	-.45	-33.74	+17.21	+118.57	+58.86	+1.86	+.13	-.77
$\underline{y} \le 1$	.0692	-1.55	-.59	-.41	-.59	+2.11	-.99	-4.64	+23.62	+17.27	+32.39	+1.18	+.08	-.42
$\underline{y} \le 2$	.2061	-.58	-.14	-.01	-.38	+1.15	-.31	+2.55	+17.96	-2.53	+15.54	+.60	+.04	-.23
$\underline{y} \le 3$	.4114	-.24	-.03	+.06	-.22	+.04	+.05	+2.59	+10.87	-5.23	+5.35	+.15	+.01	-.15
$\underline{y} \ge 5$	.3704	+.19	+.01	-.08	+.16	+.65	-.20	-1.85	-9.17	+5.29	+.22	+.22	+.01	+.22
$\underline{y} \ge 6$	.1958	+.21	+.01	-.11	+.09	+1.40	-.26	+.33	-8.38	+2.59	+9.74	+1.00	+.08	+.50
$\underline{y} \ge 7$	.0867	+.26	+.04	-.11	-.02	+1.63	-.05	+5.66	-4.44	-6.42	+27.66	+2.46	+.24	+1.07
$\underline{y} \ge 8$	.0321	+.32	+.09	-.09	+.01	+.62	+.59	+15.21	+3.51	-21.60	+59.08	+4.88	+.56	+2.15
$\underline{y} \ge 9$	.0100	+.41	+.17	-.04	+.54	-2.38	+1.82	+30.46	+16.81	-40.47	+114.02	+8.64	+1.13	+3.98
$\underline{y} \ge 10$	.0026	+.51	+.27	+.03	+2.08	-8.02	+3.79	+53.67	+37.52	-59.38	+213.40	+14.31	+2.11	+6.95
n = 100 p = .2														
$\underline{y} \le 9$	.0023	-.26	-.13	-.10	-.08	-3.25	+1.69	-28.51	-13.56	+85.65	+114.07	+3.67	+.21	-.74
$\underline{y} \le 11$	.0126	-.15	-.06	-.04	-.02	-.48	+.29	-13.42	-.55	+33.55	+70.08	+2.62	+.15	-.45
$\underline{y} \le 13$	.0469	-.09	-.03	-.01	-.02	+.35	-.10	-4.48	+5.59	+11.02	+40.96	+1.72	+.10	-.26
$\underline{y} \le 15$	.1285	-.05	-.01	+.00	-.02	+.38	-.12	-.21	+7.14	+1.39	+21.79	+.99	+.06	-.15
$\underline{y} \le 17$	.2712	-.03	-.00	+.01	-.02	+.18	-.05	+1.16	+6.21	-1.92	+9.53	+.43	+.03	-.09
$\underline{y} \le 19$	.4602	-.02	-.00	+.01	-.02	+.00	+.00	+1.08	+4.34	-2.15	+2.19	+.06	+.00	-.07
$\underline{y} \ge 22$	.3460	+.02	+.00	-.01	+.02	+.14	-.04	-.98	-4.69	+2.27	+2.99	+.28	+.02	+.12
$\underline{y} \ge 24$	.1891	+.03	+.00	-.01	+.01	+.29	-.07	-.10	-4.75	+.90	+12.39	+.90	+.06	+.24
$\underline{y} \ge 26$	.0875	+.03	+.01	-.01	+.00	+.36	-.05	+2.21	-3.40	-3.33	+28.25	+1.89	+.13	+.46
$\underline{y} \ge 28$	.0342	+.04	+.01	-.00	+.00	+.17	+.09	+6.49	-.16	-11.00	+53.67	+3.34	+.24	+.83
$\underline{y} \ge 30$	.0112	+.06	+.02	+.00	+.04	-.46	+.38	+13.34	+5.50	-22.00	+93.96	+5.34	+.40	+1.38
$\underline{y} \ge 32$	.0031	+.07	+.04	+.01	+.15	-1.75	+.89	+23.55	+14.28	-35.46	+158.55	+8.00	+.64	+2.19

TABLE 7.5. Event $\underline{y} \leq k$ or $\underline{y} \geq k+1$, exact binomial probability and relative tail error in per cent. for some normal and Poisson approximations. Case p = .05

Event		Proba-bility	Peizer-Pratt=(5.12) $\varepsilon_1=0$ $\varepsilon_2=0$	$\varepsilon_1=.02$ $\varepsilon_2=.02$	$\varepsilon_1=.02$ $\varepsilon_2=.13$	corrected square roots (5.22)	(4.22)	Camp-Paulson (4.6)	square root (3.20) b=-1/4	(3.20) b=0	binom. u_b (2.2)	Poisson approximations for λ_i Table 6.2, p.107 $\lambda_1=np$	λ_2	λ_5	λ_3
n =	5	p = .05													
$\underline{y} \geq$	1	.2262	+4.86	-.04	-1.99	+.70	+15.71	-.68	+5.10	-24.31	+34.37	-2.22	-.02	-.00	+.01
$\underline{y} \geq$	2	.0226	+4.96	+1.25	-1.52	+13.64	-1.73	+12.37	+92.01	+38.43	-77.16	+17.29	+1.20	+.14	+1.00
$\underline{y} \geq$	3	.0012	+7.79	+4.47	+.93	+114.07	-57.06	+46.53	+449.91	+300.37	-99.83	+86.64	+6.59	+1.57	+5.73
n =	20	p = .05													
$\underline{y} \leq$	0	.3585	-3.29	-1.03	-.76	-2.86	-.71	+.37	+6.66	+33.84	-15.20	+2.62	+.02	+.00	-.01
$\underline{y} \geq$	2	.2642	+1.26	+.09	-.19	+.81	+4.19	-.97	-1.44	-20.62	+15.07	+.03	+.03	+.00	+.03
$\underline{y} \geq$	3	.0755	+1.24	+.13	-.25	+.41	+4.73	+.64	+17.57	-7.84	-17.99	+6.38	+.22	+.01	+.14
$\underline{y} \geq$	4	.0159	+1.47	+.43	-.05	+3.58	-2.91	+5.67	+62.94	+25.45	-67.55	+19.41	+.64	+.04	+.41
$\underline{y} \geq$	5	.0026	+1.89	+.89	+.33	+14.78	-19.67	+15.01	+159.76	+97.52	-93.59	+42.19	+1.44	+.12	+.97
n =	100	p = .05													
$\underline{y} \leq$	0	.0059	-5.97	-2.93	-2.86	+1.15	-8.27	+2.46	-53.86	-1.67	+228.92	+13.81	+.11	+.00	-.05
$\underline{y} \leq$	1	.0371	-1.63	-.52	-.46	-.58	+1.23	-1.05	-16.22	+21.99	+46.02	+9.02	+.08	+.00	-.03
$\underline{y} \leq$	2	.1183	-.69	-.15	-.11	-.43	+1.52	-.77	-.86	+24.02	+6.27	+5.40	+.05	+.00	-.02
$\underline{y} \leq$	3	.2578	-.34	-.04	-.01	-.27	+.64	-.24	+3.29	+18.87	-4.73	+2.79	+.02	+.00	-.01
$\underline{y} \leq$	4	.4360	-.18	-.01	+.01	-.16	-.02	+.05	+3.04	+12.59	-6.13	+1.03	+.01	+.00	-.01
$\underline{y} \geq$	6	.3840	+.15	+.00	-.02	+.13	+.46	-.21	-2.60	-11.80	+6.58	+.01	+.01	+.00	+.01
$\underline{y} \geq$	7	.2340	+.17	+.01	-.02	+.11	+.94	-.32	-1.13	-12.11	+4.98	+1.64	+.04	+.00	+.02
$\underline{y} \geq$	8	.1280	+.18	+.01	-.02	+.09	+1.24	-.27	+2.69	-10.20	-1.79	+4.23	+.09	+.00	+.03
$\underline{y} \geq$	9	.0631	+.21	+.03	-.01	+.09	+1.09	+.05	+9.51	-5.61	-14.17	+7.93	+.16	+.00	+.06
$\underline{y} \geq$	10	.0282	+.23	+.05	+.00	+.20	+.26	+.73	+20.19	+2.28	-30.92	+12.91	+.25	+.01	+.09
$\underline{y} \geq$	11	.0115	+.27	+.08	+.03	+.50	-1.47	+1.85	+35.93	+14.36	-49.37	+19.38	+.38	+.01	+.15
$\underline{y} \geq$	12	.0043	+.31	+.11	+.05	+1.11	-4.26	+3.50	+58.50	+32.00	-66.55	+27.58	+.54	+.02	+.22
$\underline{y} \geq$	13	.0015	+.35	+.15	+.09	+2.16	-8.19	+5.74	+90.61	+57.31	-80.22	+37.87	+.74	+.03	+.32

normal and Poisson approximation. Fig. 7.2 also sketches a boundary between these two. For n < 10 the discrete character of the binomial distribution makes the determination of such a boundary almost impossible. Even for larger values of n there exists on both sides of the boundary a strip of parameter pairs (n,p) for which the decision is difficult.

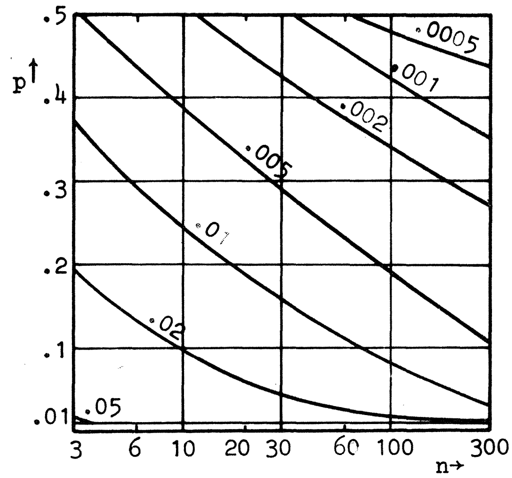

Fig. 7.1. Maximal absolute error for the square root approximation $\Phi(2(k+\frac{3}{4})^{\frac{1}{2}}q^{\frac{1}{2}} - 2(n-k-\frac{1}{4})^{\frac{1}{2}}p^{\frac{1}{2}})$.

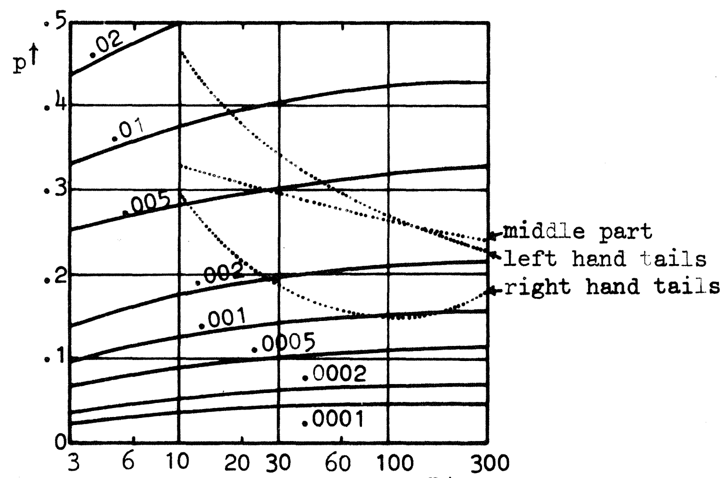

Fig. 7.2. Maximal absolute error for Bolshev's approximation $F_{\lambda_2}(k)$, $\lambda_2 = (2n-k)p/(2-p)$, and rough boundaries above which the square root approximations are better than F_{λ_2}.

CHAPTER IV: NORMAL, POISSON AND BINOMIAL APPROXIMATIONS TO THE HYPERGEOMETRIC DISTRIBUTION

1. NOTATION, EXACT VALUES, SUMMARY

Throughout this Chapter, the random variable $\underline{a}$ has a hypergeometric distribution with parameters n,r,N (positive integers). Its distribution function is

$$(1.1)\qquad H_{n,r,N}(k) = P[\underline{a} \le k] = \sum_{j=0}^{k} \binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n}.$$

This distribution is connected to the following 2×2 table with fixed marginals:

(1.2)

$\underline{a}$	$\underline{b}$	n
$\underline{c}$	$\underline{d}$	m=N-n
r	s=N-r	N

We shall always assume that the 2×2 table is arranged in such a way that $n \le r \le \frac{1}{2}N$. This is no restriction, as one may always interchange the rows and/or columns; for fixed marginals one has

$$(1.3)\qquad \begin{aligned} &P[\underline{a} \le k] = P[\underline{b} \ge n-k] = P[\underline{c} \ge r-k] = \\ &= P[\underline{d} \le N-n-r+k]; \end{aligned}$$

the random variables $\underline{b}$, $\underline{c}$ and $\underline{d}$ are fully determined by $\underline{a}$ and will not be considered any further.

Under the convention $n \le r \le s \le m$, the hypergeometric random variable $\underline{a}$ can assume the values 0, 1, ..., n. The argument k of its distribution function is an integer for which we assume that $0 \le k < n$, as trivially $H_{n,r,N}(k) = 1$ for k = n.

We shall use the notations $\mu = nrN^{-1}$ and $\tau^2 = mnrsN^{-3}$. It is well known that $E\underline{a} = \mu$ and $\sigma^2(\underline{a}) = \tau^2 N(N-1)^{-1}$. We work with τ^2 instead of $\sigma^2(\underline{a})$

because this simplifies the calculations. The marginal fractions will be denoted by a tilde: $\tilde{m} = m/N$, $\tilde{n} = n/N$, $\tilde{r} = r/N$ and $\tilde{s} = s/N$. Obviously $\tilde{m} + \tilde{n} = \tilde{r} + \tilde{s} = 1$, and $\frac{1}{4} \leq \tilde{\tilde{m}}\tilde{\tilde{s}} = \tau^2/\mu < 1$.

A 2×2 table with $r = s = \frac{1}{2}N$ will be called *symmetric*, as in this case $P[\underline{a} = k] = P[\underline{a} = n-k]$ and $H_{n,\frac{1}{2}N,N}(k) = 1 - H_{n,\frac{1}{2}N,N}(n-k-1)$. If $m = n = r = s = \frac{1}{2}N$ we shall speak of a *doubly symmetric* table. In section 2 we shall see that symmetric and doubly symmetric tables play a special role for most normal approximations.

Let us briefly review the three situations in which 2×2 tables and hypergeometric distributions are frequently used.

(i) *attributive sampling from a finite population.* Let an urn contain N balls, of which r are red and s = N-r are black. In a random sample of size n without replacement, the number $\underline{a}$ of red balls has a hypergeometric (n,r,N) distribution. It is usually applied for confidence intervals and tests of hypotheses concerning the fraction $\tilde{r} = r/N$ of red balls in the urn.

(ii) *homogeneity problem: comparison of two binomial parameters.* Let n independent experiments with unknown success probability p_1 lead to $\underline{a}$ successes, whereas m = N-n independent experiments with unknown success probability p_2 produce $\underline{c}$ successes. A test for the hypothesis $p_1 = p_2$, based on the observed values $\underline{a} = a$ and $\underline{c} = c$, uses the hypergeometric (n,r,N) distribution of $\underline{a}$ under the condition $\underline{a}+\underline{c} = a+c = r$. This test procedure is uniformly most powerful unbiased (see e.g. LEHMANN, 1959, section 4.5). We shall not review the numerous discussions on the validity of such a conditional test, nor shall we go into the point that its actual significance level can be much lower than its nominal one when randomization is considered to be unacceptable (BOSCHLOO, 1970).

(iii) *double dichotomy problem: dependence of two properties.* Let X and Y be two properties that any individual element of a given population may or may not possess. In a random sample of size N from this population, let $\underline{a}$ elements have both properties, $\underline{b}$ have X but not Y, $\underline{c}$ have Y but not X, and $\underline{d} = N-\underline{a}-\underline{b}-\underline{c}$ have neither. A test for the statistical independence of X and Y, based on the observed values $\underline{a} = a$,

$\underline{b}$ = b, $\underline{c}$ = c and $\underline{d}$ = d, uses the hypergeometric (n,r,N) distribution of $\underline{a}$ under the conditions $\underline{a}+\underline{b}$ = a+b = n and $\underline{a}+\underline{c}$ = a+c = r. Again the conditional test is uniformly most powerful unbiased, and again its validity has been disputed. We refer to LEHMANN (1959, section 4.6), who also mentions improved designs for situations where the marginal probabilities of X and Y are known.

Tables of $H_{n,r,N}(k)$ exist only for rather limited ranges of the four entries k,n,r,N. LIEBERMAN & OWEN (1961) give the distribution function and the individual terms for:

$N = 2(1)25$, $1 \leq n \leq r \leq N-1$, $0 \leq k \leq n$;
$N = 26(1)50$, $1 \leq n \leq r \leq \frac{1}{2}N$, $0 \leq k \leq n$;
$N = 60(10)100$, $1 \leq n \leq r \leq \frac{1}{2}N$, $0 \leq k \leq n$;
$N = 1000$, $r=500$, $1 \leq n \leq 500$, $0 \leq k \leq \frac{1}{2}n$, $H > 10^{-6}$;
$N = 100(100)2000$, $r = \frac{1}{2}N$, $n = r$ or $n = r-1$, $0 \leq k \leq \frac{1}{2}n$, $H > 10^{-6}$.

A more complete tabulation would ask for even more than the present 600 pages of the table. Interpolation is rather cumbersome and not very accurate.

Suppose that $N \to \infty$, and that the marginals m,n,r,s may vary with N, but such that always $n \leq r \leq s \leq m$. As pointed out by VAN EEDEN & RUNNENBURG (1960), there exists a limiting distribution of $\underline{a}$ if and only if $\mu = nrN^{-1}$ and $\tau^2 = mnrsN^{-3}$ tend to (finite or infinite) limits. This limiting distribution is:

(a) normal (after standardization) if and only if $\mu \to \infty$ and $\tau^2 \to \infty$;
(b) binomial $(\mu_0^2/(\mu_0-\tau_0^2), (\mu_0-\tau_0^2)/\mu_0)$ if and only if $\mu \to \mu_0$ and $\tau^2 \to \tau_0^2$ where $0 < \tau_0^2 < \mu_0 < \infty$; we shall see in section 4 that the assumptions imply that the value of n must be constant for sufficiently large N; this constant value equals $\mu_0^2/(\mu_0-\tau_0^2)$, which is thus always an integer;
(c) Poisson (μ_0) if and only if $\mu \to \mu_0$ and $\tau^2 \to \mu_0$, where $0 < \mu_0 < \infty$;
(d) degenerate if and only if $\mu \to 0$ and $\tau \to 0$.

Approximations based on (a), (b) and (c) will be discussed in sections 2, 4 and 3 respectively. Section 5 gives numerical information about errors and a survey of the results.

2. NORMAL AND χ^2 APPROXIMATIONS

2a. INTRODUCTION

When all marginals of a 2×2 table are large (say: when the expected value for each cell exceeds 5), it is often advised to use the classical normal approximation to the hypergeometric distribution, with continuity correction $\frac{1}{2}$:

$$H_{n,r,N}(k) = P[\underline{a} \leq k] \simeq \Phi\left(\frac{k+\frac{1}{2}-nrN^{-1}}{\{mnrsN^{-2}(N-1)^{-1}\}^{\frac{1}{2}}}\right). \tag{2.1}$$

Also for large marginals, one takes for the two-sided probability of exceedance, in the testing problems described in section 1, the probability that a χ_1^2 random variable exceeds

$$\frac{\{|ad-bc| - \frac{1}{2}N\}^2 N}{mnrs} = \frac{\{|a-nrN^{-1}| - \frac{1}{2}\}^2}{mnrsN^{-3}}. \tag{2.2}$$

The equality of the two expressions in (2.2) follows immediately from $d = N-n-r+a$, $b = n-a$, $c = r-a$. As χ_1^2 is the square of standard normal variable, this χ^2 approximation is for $|a-nrN^{-1}| \geq \frac{1}{2}$ equivalent to the normal approximation

$$H_{n,r,N}(k) \simeq \Phi(\chi), \text{ where } \chi = \frac{k+\frac{1}{2}-nrN^{-1}}{\{mnrsN^{-3}\}^{\frac{1}{2}}} = \frac{k+\frac{1}{2}-\mu}{\tau}. \tag{2.3}$$

Subsection 2b presents an asymptotic expansion for the exact normal deviate $\xi = \xi(k,n,r,N)$ defined by $\Phi(\xi) = H_{n,r,N}(k)$. The explicit solution of ξ from this transcendental equation is of course impossible, but we shall use the expansion for ξ in order to compare the asymptotic errors of some simple functions of k,n,r,N that are approximations to ξ. Examples of such functions are given in (2.1) and (2.3). A surprisingly simple and relatively accurate normal approximation will be derived in subsection 2c. Asymptotic and numerical conclusions about errors will be given in subsection 2d. Subsection 2e discusses more accurate normal approximations.

For all asymptotic expansions in section 2, it is tacitly assumed that

N, $\mu = nrN^{-1}$ and $\tau^2 = mnrsN^{-3}$ all tend to infinity, and that the exact normal deviate ξ is bounded. The latter condition implies that the values of the hypergeometric distribution function are bounded away from 0 and 1. This is no serious restriction (cf. section II.2). It enables us to consider O- and o-symbols as uniform in ξ and k. The assumed convergence of N, μ and τ implies that all marginals m,n,r,s tend to infinity. From our convention $n \leq r \leq \frac{1}{2}N$ it follows that the fractional marginals $\tilde{m}$, $\tilde{n}$, $\tilde{r}$, $\tilde{s}$, satisfy

$$(2.4) \qquad 0 < \tilde{n} \leq \tilde{r} \leq \tfrac{1}{2} \leq \tilde{s} \leq \tilde{m} < 1 \qquad\qquad (\tilde{n} = nN^{-1}, \text{ etc}).$$

It may happen that $\tilde{n}$, possibly even $\tilde{n}$ and $\tilde{r}$, tend to zero, but if they do, the convergence must be slow enough to be compatible with our assumption that $\mu = nrN^{-1} = \widetilde{nr}N$ tends to infinity. Recall that for normal approximations to the binomial distribution we have assumed that $npq \to \infty$, but possibly $p \to 0$ (sections III.1 and III.2).

As $\tau^2 = \mu\widetilde{ms}$, with $\tilde{m}$ and $\tilde{s}$ between $\frac{1}{2}$ and 1, it is clear that μ and τ^2 are asymptotically of the same order of magnitude. N is also of this order whenever $\tilde{n}$ is bounded away from zero.

2b. THE EXACT NORMAL DEVIATE

THEOREM

The exact normal deviate ξ defined by $\Phi(\xi) = H_{n,r,N}(k)$, satisfies the the expansion

$$(2.5) \qquad \begin{aligned} \xi &= \chi + \\ &+ \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(1-\chi^2)/6 + \\ &+ \tau^{-2}\{\chi^3(5-14\widetilde{mn}-14\widetilde{rs}+38\widetilde{mnrs}) + \chi(-2+2\widetilde{mn}+2\widetilde{rs}+10\widetilde{mnrs})\}/72 + \\ &+ o(\tau^{-2}), \end{aligned}$$

where

$$(2.6) \qquad \begin{aligned} &\chi = (k+\tfrac{1}{2}-\mu)\tau^{-1}; \; \mu = nrN^{-1}; \; \tau^2 = mnrsN^{-3}; \\ &\tilde{m} = mN^{-1}, \; \tilde{n} = nN^{-1}, \; \tilde{r} = rN^{-1}, \; \tilde{s} = sN^{-1}. \end{aligned}$$

PROOF. As we have assumed that $\tau \to \infty$, we may suppose that $\log \tau > 0$. Let h denote the smallest integer exceeding $\mu - \tau \log \tau$. Then $|h-\mu| < \tau \log \tau$. On the other hand the exact normal deviate ξ is assumed to be bounded, and it is easy to show that ξ, χ and $(k-\mu)\tau^{-1}$ are of the same order of magnitude. Thus $h \leq k$, and for $h \leq j \leq k$ we have $|j-\mu| < \tau \log \tau$. The proof is based on an asymptotic expansion for

$$(2.7) \qquad H_{n,r,N}(k) = \sum_{j=0}^{h} P[\underline{a}=j] + \sum_{j=h+1}^{k} P[\underline{a}=j],$$

and consists of four steps:

step 1) expansion of $P[\underline{a}=j]$ for all j satisfying $h \leq j \leq k$;
step 2) derivation of an upper bound for the first sum in (2.7);
step 3) evaluation of the second sum, replacing sums by integrals;
step 4) combination of the results into a proof of (2.5).

Step 1). Let $h \leq j \leq k$. Consider the j-th term of the hypergeometric distribution:

$$(2.8) \qquad P[\underline{a}=j] = \binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n} = \frac{m!n!r!s!}{N!j!(n-j)!(r-j)!(s-n+j)!} .$$

We introduce the notation $t_j = (j-\mu)/\tau$; then $|t_j| < \log \tau$ (see above). From $j = \mu + t_j\tau$ follows $n-j = n - nrN^{-1} - t_j\tau = nsN^{-1} - t_j\tau$, and similar expressions for $r-j$ and $s-n+j$; one can see that they all tend to infinity. Now apply Stirling's formula to each factorial in (2.8), e.g.

$$(2.9) \qquad \log j! = (j+\tfrac{1}{2}) \log j - j + \tfrac{1}{2} \log 2\pi - 1/(12j) + O(j^{-3}).$$

Because $j = \mu + t_j\tau$, the term $(j+\frac{1}{2}) \log j$ becomes

$$(2.10) \qquad (j+\tfrac{1}{2}) \log \mu + (\mu + t_j\tau + \tfrac{1}{2}) \log (1 + t_j\tau\mu^{-1}),$$

and the second logarithm can be expanded because $t_j\tau\mu^{-1} = O(\tau^{-1} \log \tau)$. Proceeding in a similar way with the other factorials, one finds an expression for $\log P[\underline{a}=j]$. After simplification involving relations such as

$m+n = r+s = N$ and (2.6), the result can be written in the form

$$
\begin{aligned}
P[\underline{a}=j] &= \phi(t_j)\Big[\tau^{-1} + \tau^{-2}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(t_j^3-3t_j)/6 + \\
(2.11) \qquad &+ \tau^{-3}\{-(1-\widetilde{mn})(1-\widetilde{rs})/12 + (1-2\widetilde{mn})(1-2\widetilde{rs})t_j^2/4-(1-3\widetilde{mn})(1-3\widetilde{rs})t_j^4/12 + \\
&+ (1-4\widetilde{mn})(1-4\widetilde{rs})(t_j^6-6t_j^4+9t_j^2)/72\} + o(\tau^{-3})\Big];
\end{aligned}
$$

we recall that ϕ denotes the standard normal density, that $t_j = (j-\mu)\tau^{-1}$ and that the other symbols are explained in (2.6) and in section 1. By explicit consideration of the contributions to the remainder term, one can establish that (2.11) holds with a uniform o-symbol for $h \leq j \leq k$.

Step 2). For the derivation of an upper bound for the first sum in (2.7), we observe that for any $j \leq h-1$ one finds from (2.8) that

$$
(2.12) \qquad \frac{P[\underline{a}=j]}{P[\underline{a}=j+1]} = \frac{(j+1)(s-n+j+1)}{(n-j)(r-j)} \leq \frac{h(s-n+h)}{(n-h+1)(r-h+1)} .
$$

Now use $h = \mu + t_h\tau$; a little calculation shows that

$$
(2.13) \qquad \frac{h(s-n+h)}{(n-h+1)(r-h+1)} = 1 + t_h\tau^{-1} + O(t_h^2\tau^{-2}).
$$

This lies between 0 and 1, at least for large values of τ, as the definition of h at the beginning of the proof implies that $t_h \simeq -\log\tau$. Now it is clear from (2.12) that

$$
(2.14) \qquad \sum_{j=0}^{h} P[\underline{a}=j] \leq P[\underline{a}=h] \times \sum_{j=0}^{h} \left\{\frac{h(s-n+h)}{(n-h+1)(r-h+1)}\right\}^j ;
$$

the first factor equals $\phi(t_h)\{\tau^{-1} + O(t_h^3\tau^{-2})\}$ because of (2.11), and the second factor is majorized by the infinite geometric series, which equals $-\tau t_h^{-1} + O(1)$ because of (2.13). Thus the right hand side of (2.14) has a leading term $-t_h^{-1}\phi(t_h)$. From $t_h \simeq -\log\tau$ it follows that $\phi(t_h) =$ $= O(\tau^{-\frac{1}{2}\log\tau})$, so the first sum of (2.7) is certainly $o(\tau^{-2})$.

Step 3). When (2.11) is applied to each term of the second sum in (2.7), one encounters expressions of the form

$$Z_i = \tau^{-1} \sum_{j=h+1}^{k} t_j^i \, \phi(t_j) \qquad \text{(integer } i \geq 0\text{)}. \tag{2.15}$$

For their reduction, similar to Euler's summation formula, we introduce

$$L_i(z) = \int_{-\infty}^{z} w^i \, \phi(w) \, dw. \tag{2.16}$$

Then, uniformly in j for $h+1 \leq j \leq k$,

$$\int_{t_j-\frac{1}{2}\tau^{-1}}^{t_j+\frac{1}{2}\tau^{-1}} w^i \, \phi(w) \, dw = L_i(t_j+\tfrac{1}{2}\tau^{-1}) - L_i(t_j-\tfrac{1}{2}\tau^{-1}) =$$

$$= \phi(t_j)\left[t_j^i/\tau + \{i(i-1)t_j^{i-2} - (2i+1)t_j^i + t_j^{i+2}\}/(24\tau^3)\right] + \tag{2.17}$$

$$+ O(t_j^{i+4} \, \phi(t_j) \, \tau^{-5}),$$

because the even powers of τ^{-1} cancel in the Taylor expansions of $L_i(t_j \pm \frac{1}{2}\tau^{-1})$, and for L_i', L_i''' and the leading term of the fifth derivative one can deduce expressions from (2.16), using $\phi'(z) = - z \, \phi(z)$. It follows from (2.17) that

$$\begin{aligned} L_i(\chi) - L_i(t_h+\tfrac{1}{2}\tau^{-1}) &= Z_i + \{i(i-1)Z_{i-2} + \\ &- (2i-1) \, Z_i + Z_{i+2}\}/(24\tau^2) + o(\tau^{-3}). \end{aligned} \tag{2.18}$$

Now application of (2.11) to each term in the second sum of (2.7) leads to an expression involving the sums Z_6, Z_4, Z_3, Z_2, Z_1 and Z_0. By (2.18), they can be recurrently expressed in the integrals L_i, with an error $o(\tau^{-3})$. Next observe that partial integration gives

$$L_i(z) = -L_{i-1}'(z) + (i-1)L_{i-2}(z) \quad (i \geq 2). \tag{2.19}$$

As obviously $L'_{i-1}(z) = z^{i-1}\phi(z)$, one obtains that

$$
(2.20) \qquad
\begin{aligned}
L_0(z) &= \Phi(z), \\
L_1(z) &= -\phi(z), \\
L_2(z) &= -z\phi(z) + \Phi(z), \\
L_3(z) &= -z^2\phi(z) - 2\phi(z), \\
L_4(z) &= -z^3\phi(z) - 3z\phi(z) + 3\Phi(z), \\
L_6(z) &= -z^5\phi(z) - 5z^3\phi(z) - 15z\phi(z) + 15\Phi(z).
\end{aligned}
$$

Now some calculation shows that the second sum of (2.7) is equal to $W(\chi) - W(t_h+\frac{1}{2}\tau^{-1})$, where

$$
\begin{aligned}
W(z) = \Phi(z) &+ \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(1-z^2)\phi(z)/6 + \\
&+ \tau^{-2}\{z^5(-1+4\widetilde{mn}+4\widetilde{rs}-16\widetilde{mnrs}) + z^3(7-22\widetilde{mn}-22\widetilde{rs}+70\widetilde{mnrs}) + \\
&+ z(-3+6\widetilde{mn}+6\widetilde{rs}-6\widetilde{mnrs})\}\,\phi(z)/72 + o(\tau^{-2}).
\end{aligned}
$$

As $t_h \simeq -\log\tau$, we have $W(t_h+\frac{1}{2}\tau^{-1}) = o(\tau^{-2})$, and thus the second sum of (2.7) equals $W(\chi)$.

Step 4). The hypergeometric distribution function has been split up into two sums; the first has been shown to be $o(\tau^{-2})$ and the second is $W(\chi)$. In the usual way (cf. HILL & DAVIS, 1968) one now deduces (2.5) from $\Phi(\xi) = H_{n,r,N}(k) = W(\chi)$. This completes the proof of the theorem.

Straightforward inversion of (2.5) shows that

$$
(2.22) \qquad
\begin{aligned}
\chi = (k+\tfrac{1}{2}-nrN^{-1})\,(mnrsN^{-3})^{-\frac{1}{2}} &= \xi + \\
&+ \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(\xi^2-1)/6 + \\
&+ \tau^{-2}\{\xi^3(-1-2\widetilde{mn}-2\widetilde{rs}+26\widetilde{mnrs}) + \xi(-2+14\widetilde{mn}+14\widetilde{rs}-74\widetilde{mnrs})\}/72 + \\
&+ o(\tau^{-2}).
\end{aligned}
$$

2c. SQUARE ROOT APPROXIMATIONS

One can derive a transformation of the hypergeometric random variable $\underline{a}$ which makes its variance asymptotically constant (cf. section III.3c). This leads to the consideration of normal deviates like

$$(2.23)\qquad 2(n+\delta_1)^{\frac{1}{2}}(N+\delta_2)^{\frac{1}{2}}(m+\varepsilon)^{-\frac{1}{2}}\left[\arcsin\left\{\frac{k+\frac{1}{2}+\beta}{n+\gamma}\right\}^{\frac{1}{2}} - \arcsin \tilde{r}^{\frac{1}{2}}\right].$$

Here β, γ, δ_1, δ_2 and ε are arbitrary constants, or possibly simple functions of the parameters n,r,N. We recall that $\tilde{r} = r/N$. Just as in the binomial case, one may replace the small angle displayed between square brackets by its sine, and consider

$$(2.24)\qquad 2(m+\varepsilon)^{-\frac{1}{2}}\left[(k+1)^{\frac{1}{2}}s^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}r^{\frac{1}{2}}\right];$$

we have chosen $\beta = \frac{1}{2}$, $\gamma = \delta_1 = 1$ and $\delta_2 = 0$ because the square brackets in (2.24) now contain only square roots of integers, easily found in tables. The leading term of the error of (2.24), not derived here, turns out to be proportional to $\xi^2 - 4$. It would be attractive if also the following term were small for $\xi = \pm 2$: in many applications τ is not large enough to make this next term negligible, and for symmetric tables ($r = s = \frac{1}{2}N$) it is even the leading term. It turns out that the choice $\varepsilon = -3N/(8n)$ achieves our goal for the doubly symmetric case, and is also reasonably good for most other cases. For this value the expansion of (2.24), not explicitly derived here, is

$$(2.25)\qquad \begin{aligned} &2\{m - 3N/(8n)\}^{-\frac{1}{2}}\left[(k+1)^{\frac{1}{2}}s^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}r^{\frac{1}{2}}\right] = \xi + \\ &+ \tau^{-1}(2-\tilde{m})(\tilde{s}-\tilde{r})(4 - \xi^2)/12 + \\ &+ \tau^{-2}\{\xi^3(-1+4\tilde{m}-\tilde{m}^2-2\widetilde{rs}+2\widetilde{mrs}-5\tilde{m}^2\widetilde{rs}) + \\ &+ \xi(-2-10\tilde{m}-2\tilde{m}^2+27\tfrac{1}{2}\widetilde{rs}-14\widetilde{mrs}+26\tilde{m}^2\widetilde{rs})\}/72 + \\ &+ o(\tau^{-2}). \end{aligned}$$

The use of (2.24) with $\varepsilon = -\,3N/(8n)$ has some disadvantages. It is nearly always more accurate than the classical normal or χ^2 approximation for probabilities between .01 and .05 or between .95 and .99, but it is rather inaccurate in the middle part of the distribution, and sometimes also for small tails. Besides, calculation of $\{m - 3N/(8n)\}^{-\frac{1}{2}}$ is unattractive. Furthermore, it seems impossible to find a simple modification of (2.24) which is optimal for other values than $|\xi| = 2$.

As explained in subsection 2d, some of these drawbacks would be eliminated if we could replace the factor $(2-\tilde{m})$ in (2.25) by $(\tilde{m}-\tilde{n})$. Motivated by a remark of WISE (1954) concerning binomial approximations to the hypergeometric distribution (section 4), we have replaced in (2.24) r by $r-\frac{1}{2}k$ and s by $s-\frac{1}{2}n+\frac{1}{2}k$: in the terminology of model (i) of section 1, the number of red balls before the drawing of n balls was replaced by the average number of red balls during an experiment ending up with k red balls drawn, and similarly for the black balls. Expansion of

$$(2.26)\qquad (k+1)^{\frac{1}{2}}(s-\tfrac{1}{2}n+\tfrac{1}{2}k)^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(r-\tfrac{1}{2}k)^{\frac{1}{2}}$$

shows that it would be still better to take the number of balls at the end of the experiment, i.e. $s-n+k$ and $r-k$. A little calculation now leads to the proposal

$$(2.27)\qquad 2(N+\delta)^{-\frac{1}{2}}\left[(k+\tfrac{1}{2}+\beta)^{\frac{1}{2}}(s-n+k+\tfrac{1}{2}+\beta)^{\frac{1}{2}} - (n-k-\tfrac{1}{2}+\beta)^{\frac{1}{2}}(r-k-\tfrac{1}{2}+\beta)^{\frac{1}{2}}\right],$$

where β and δ are arbitrary constants. Because of (2.3) we may substitute $k+\frac{1}{2} = nrN^{-1}+\chi\tau$. We find that

$$(2.28)\qquad \begin{aligned}(k+\tfrac{1}{2}+\beta)(s-n+k+\tfrac{1}{2}+\beta) &= (nrN^{-1}+\chi\tau+\beta)(msN^{-1}+\chi\tau+\beta) = \\ &= \tau^2N\{1 + \frac{(ms+nr)\chi}{\tau N^2} + \frac{\chi^2}{N} + \frac{(ms+nr)\beta}{\tau^2N^2} + \frac{2\beta\chi}{\tau N} + \frac{\beta^2}{\tau^2N}\}.\end{aligned}$$

A similar expression can be found for $(n-k-\frac{1}{2}+\beta)(r-k-\frac{1}{2}+\beta)$. Next we take square roots, and use

$$(2.29)\qquad (1+R)^{\frac{1}{2}} = 1 + R/2 - R^2/8 + R^3/16 + O(R^4).$$

It follows after some calculations, including the use of $\tau^2 = mnrsN^{-3}$ and $m+n = r+s = N$ in various ways, that the deviate (2.27) has the expansion

$$\chi + \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(12\beta - 3\chi^2)/12 +$$

(2.30)
$$+ \tau^{-2}\{\chi^3(9-27\widetilde{mn}-27\widetilde{rs}+72\widetilde{mnrs}) + \chi(-36\beta+72\beta\widetilde{mn}+72\beta\widetilde{rs}-36\delta\widetilde{mnrs})\}/72 +$$

$$+ o(\tau^{-2}).$$

Substitution of the expansion (2.22) of χ in terms of ξ leads to

$$2(N+\delta)^{-\frac{1}{2}}\left[(k+\tfrac{1}{2}+\beta)^{\frac{1}{2}}(s-n+k+\tfrac{1}{2}+\beta)^{\frac{1}{2}} - (n-k-\tfrac{1}{2}+\beta)^{\frac{1}{2}}(r-k-\tfrac{1}{2}+\beta)^{\frac{1}{2}}\right] =$$

$$= \xi + \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(12\beta-2 - \xi^2)/12 +$$

(2.31)
$$+ \tau^{-2}\{\xi^3(2-5\widetilde{mn}-5\widetilde{rs}+2\widetilde{mnrs}) + \xi(4-36\beta-10\widetilde{mn}+72\beta\widetilde{mn}+$$

$$-10\widetilde{rs}+72\beta\widetilde{rs}+22\widetilde{mnrs}-36\delta\widetilde{mnrs})\}/72 +$$

$$+ o(\tau^{-2}).$$

The choice $\beta = \frac{1}{2}$, $\delta = -1$ leads to the important special case

$$2(N-1)^{-\frac{1}{2}}\left[(k+1)^{\frac{1}{2}}(s-n+k+1)^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(r-k)^{\frac{1}{2}}\right] =$$

$$= \xi + \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(4 - \xi^2)/12 +$$

(2.32)
$$+ \tau^{-2}\{\xi^3(2-5\widetilde{mn}-5\widetilde{rs}+2\widetilde{mnrs}) + \xi(-14+26\widetilde{mn}+26\widetilde{rs}+58\widetilde{mnrs})\}/72 +$$

$$+ o(\tau^{-2}),$$

which is attractive as it contains only square roots of integers. We shall see in the subsection 2d that it is generally the best simple normal approximation.

<u>EXAMPLE</u>. Evaluation of $\mathbf{P}[\underline{a} \leq 11]$ in the 2×2 table $\begin{smallmatrix} 11 & 39 \\ 43 & 68 \end{smallmatrix}$. One has $N = 11+39+43+68 = 161$; for the deviate (2.32) one obtains

$$(2.33) \qquad 2(161-1)^{-\frac{1}{2}}\{(11+1)^{\frac{1}{2}}(68+1)^{\frac{1}{2}} - 39^{\frac{1}{2}}43^{\frac{1}{2}}\} = 2 \times 160^{-\frac{1}{2}}\{828^{\frac{1}{2}}-1677^{\frac{1}{2}}\},$$

which is -1.9252. Thus the answer according to (2.32) is $\Phi(-1.9252)=.0271$; the exact value is .0269. The classical χ^2 and normal approximations lead to .0286 and .0290 respectively. Although the difference in accuracy is not always so spectacular, (2.32) is nearly always better for tails of less than .06 and frequently also for other values.

2d. COMPARISON OF SIMPLE NORMAL APPROXIMATIONS

In the literature one finds the classical normal approximation (2.1), the classical χ^2 approximation (equivalent to (2.3) unless $|k-\mu| < \frac{1}{2}$), a few arcsin deviates of type (2.23), and a somewhat complicated formula for an improved normal deviate (NICHOLSON, 1956). We shall present asymptotic and numerical considerations about the normal, χ^2 and square root approximations. NICHOLSON's results will be discussed in subsection 2e. The arcsin type will not be considered, as its calculation is cumbersome and its accuracy is no better than the accuracy of other simple deviates.

We shall first compare the normal and χ^2 approximations. A numerical investigation for $10 \leq N \leq 35$ has led HEMELRIJK (1967) to a simple footrule: for probabilities of less than .07 or more than .93, the normal is always better when $n+r \leq \frac{1}{2}N$, and the chi-squared is mostly better otherwise. We recall that any 2×2 table has been arranged in such a way that $n \leq r \leq \frac{1}{2}N$.

Let us compare this footrule to the results of the asymptotic expansions. We shall suppose that $|\xi| > 1.5$, corresponding to HEMELRIJK's restriction to tails of less than .07. In (2.22) an expansion has been given for the deviate χ which is equivalent to the χ^2 approximation (the exception for $|k-\mu| < \frac{1}{2}$ is hardly interesting, and can be completely ignored now that we assume that $|\xi| > 1.5$). If the classical normal deviate is denoted by u_h, it is obvious that

$$(2.34) \qquad u_h = \frac{k+\frac{1}{2}-\mu}{\{\frac{mnrs}{N^2(N-1)}\}^{\frac{1}{2}}} = \chi\,\{\frac{N-1}{N}\}^{\frac{1}{2}} = \chi\,\{1 - \frac{1}{2N} + O(N^{-2})\}.$$

Thus the expansion for u_h is found from (2.22) by addition of $-\tau^{-2}\xi\widetilde{mnrs}/2$. If we rewrite (2.22) as

$$(2.35) \qquad \chi = \xi + T_1 + T_{2,\chi} + o(\tau^{-2}),$$

where T_1 and $T_{2,\chi}$ denote the terms of order τ^{-1} and τ^{-2} respectively, then

$$(2.36) \qquad u_h = \xi + T_1 + T_{2,h} + o(\tau^{-2}),$$

where $T_{2,h} = T_{2,\chi} - \tau^{-2}\xi\widetilde{mnrs}/2$. It is not difficult to find from (2.22) that $T_1 \geq 0$ for any $|\xi| > 1$, whereas $T_{2,h} < T_{2,\chi} < 0$ at least for any $\xi > 1.5$, and $0 < T_{2,\chi} < T_{2,h}$ at least for any $\xi < -1.5$ (use that $0 < \widetilde{mn} \leq \widetilde{rs} \leq \frac{1}{4}$).

Now let τ be fixed, and large enough to make the $o(\tau^{-2})$ terms in (2.35) and (2.36) negligible. We shall consider two cases.

(i) *Let* $(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})$ *be so large that* $T_1 > |T_{2,h}| > |T_{2,\chi}|$. Then the normal approximation is better than the chi-squared for $\xi > 1.5$ (right hand tails): the negative contribution of $T_{2,h}$ gives more compensation for the positive error present in T_1 (although not enough). For $\xi < -1.5$ (left hand tails) the chi-squared is better: T_1 and T_2 are both positive, and $T_{2,\chi}$ adds less to the error present in T_1 than $T_{2,h}$.

(ii) *Let* $(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})$ *be so small that* $|T_{2,h}| > |T_{2,\chi}| > T_1$. Now χ^2 is better in both tails (giving less overcompensation for $\xi > 1.5$, and adding less to the error present in T_1 for $\xi < -1.5$). The difference between the two approximations becomes more important if $(\tilde{s}-\tilde{r})$ approaches 0. In the extreme case $r = s$, substitution of $\tilde{r} = \tilde{s} = \frac{1}{2}$ shows that $T_1 = 0$, $T_{2,\chi} = \tau^{-2}(-\xi^3+\xi)(1-3\widetilde{mn})/96$ and $T_{2,h} = T_{2,\chi} - \tau^{-2}\xi\widetilde{mn}/8$. This implies $|T_{2,h}| > |T_{2,\chi}|$ unless $\widetilde{mn} < 1/9$ and $|\xi|$ much smaller than 1. For $|\xi| > 1.5$ the χ^2 approximation will be considerably better; as $T_1 = 0$, it is now the leading term of the error which is smaller, not the second term.

We shall now compare these considerations to HEMELRIJK's (1967) foot-rule. Numerical investigation shows that for $n+r \leq \frac{1}{2}N$, $10 \leq N \leq 35$ and $|\xi| > 1.5$ one has nearly always situation (i). However, the distribution is so skew that even $P[\underline{a} = 0]$ corresponds to a value $\xi > -1.5$ (there are only two exceptions, for N=34 and N=35). Apart from these two exceptions, small left hand tails simply do not exist. Therefore the normal approximation is indeed better than χ^2 for $|\xi| > 1.5$, $n+r \leq \frac{1}{2}N$ and $10 \leq N \leq 35$. On the other hand situation (ii) will usually be found for $n+r > \frac{1}{2}N$ and $10 \leq N \leq 35$, and the χ^2 approximation is then indeed better.

The general situation is roughly as follows:

- for $r \simeq s$, χ^2 is much better than normal;
- for $r << s$ and left hand tails, both are very bad, but χ^2 is slightly better;
- for $r << s$ and right hand tails, both are rather bad, but χ^2 is slightly worse.

The boundary between $r \simeq s$ and $r << s$ could be something like $(s-r)N^{-\frac{1}{2}}$ = constant, but it depends also on m (or n) and k (or ξ). Anyhow, for small N more cases behave like $r \simeq s$, for large N more cases behave like $r << s$. It is indeed obvious that situation (ii) will be found for small values of τ even when e.g. $\tilde{r} = .25$, whereas for large values of τ a case with e.g. $\tilde{r} = .4$ should be considered as skew, not as almost symmetric.

Our conclusion, in agreement with a remark by ORD (1968), is that little is lost by application of the χ^2 approximation throughout: compared to the normal it is easier, frequently much better, and hardly ever much worse. Numerical values of errors (section 5) may serve to illustrate this.

For a more general discussion of simple normal approximations to the hypergeometric distribution function, we have collected some leading terms of errors in Table 2.1. The deviates χ, u_h, (2.24) and (2.27) have already been discussed, and their asymptotic expansions have been given. The deviates χ_t and u_t are modifications of χ and u_h, derived by adding to the continuity correction of $\frac{1}{2}$ an extra correction γ in order to make them asymptotically optimal at $\xi = \pm 2$ for any $r \neq s$. It follows immediately from the expansions for χ and u_h that this means $\gamma = -(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})/2$, and this leads directly to χ_t and u_t. The continuity correction $\tilde{n}+\tilde{r}-2\tilde{n}\tilde{r}$ corresponds directly to p in the binomial case, which was also optimal at $\xi = \pm 2$ ($\tilde{n} \to 0$, $\tilde{r} \to p$, cf.III. 3a).

Table 2.1 indicates that for the skew cases ($r < s$; for small N we might say $r << s$), it is best to use (2.27), with some suitable β and δ. Its error is a factor 2 smaller than the error of χ, χ_t, u_h and u_t; when (2.27) with

TABLE 2.1

Some simple normal deviates, and leading terms of their errors

deviate	error for r < s	error for r = s
$\chi = (k+\frac{1}{2}-\mu)\tau^{-1}$	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(\xi^2-1)/6$	$\tau^{-2}\{\xi^3(-6+18\widetilde{mn}) + \xi(6-18\widetilde{mn})\}/288$
$u_h = \chi(N-1)^{\frac{1}{2}}N^{-\frac{1}{2}}$	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(\xi^2-1)/6$	$\tau^{-2}\{\xi^3(-6+18\widetilde{mn}) + \xi(6-54\widetilde{mn})\}/288$
(2.27) for $\beta = \frac{1}{4}$	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(1-\xi^2)/12$	$\tau^{-2}\{\xi^3(3-18\widetilde{mn}) + \xi(-12+54\widetilde{mn}-36\delta\widetilde{mn})\}/288$
$\chi_t = (k+\tilde{n}+\tilde{r}-2\widetilde{nr}-\mu)\tau^{-1}$	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(\xi^2-4)/6$	$\tau^{-2}\{\xi^3(-6+18\widetilde{mn}) + \xi(6-18\widetilde{mn})\}/288$
$u_t = \chi_t(N-1)^{\frac{1}{2}}N^{-\frac{1}{2}}$	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(\xi^2-4)/6$	$\tau^{-2}\{\xi^3(-6+18\widetilde{mn}) + \xi(6-54\widetilde{mn})\}/288$
(2.24) for $\varepsilon = -3N/(8n)$	$\tau^{-1}(2-\tilde{m})(\tilde{s}-\tilde{r})(4-\xi^2)/12$	$\tau^{-2}\{\xi^3(-6+18\tilde{m}-9\tilde{m}^2) + \xi(19\frac{1}{2}-54\tilde{m}+18\tilde{m}^2)\}/288$
(2.27) for $\beta = \frac{1}{2}$	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(4-\xi^2)/12$	$\tau^{-2}\{\xi^3(3-18\widetilde{mn}) + \xi(-30+126\widetilde{mn}-36\delta\widetilde{mn})\}/288$
(2.27), general β and δ	$\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(12\beta-2-\xi^2)/12$	$\tau^{-2}\{\xi^3(3-18\widetilde{mn}) +$ $+ \xi(6-72\beta-18\widetilde{mn}+288\beta\widetilde{mn}-36\delta\widetilde{mn})\}/288$

$\beta = \frac{1}{2}$ is compared to (2.24), the factor $(\tilde{m}-\tilde{n})$ is never larger than $(2-\tilde{m})$, equality occurring only in the binomial limiting situation $\tilde{n} \to 0$, $\tilde{m} \to 1$. If it is desired to make (2.27) especially accurate near the prescribed values α and/or $1-\alpha$ of the distribution function, one may choose β in such a way that the leading term of the error vanishes at $\xi = \pm\xi_\alpha$, where ξ_α is the standard normal upper α fractile, defined by $\Phi(\xi_\alpha) = 1-\alpha$. Table 2.1 shows that one must take $\beta = (\xi_\alpha^2+2)/12$ for skew cases. Note that $\xi_\alpha = 2$ ($\alpha \simeq .023$) leads to $\beta = \frac{1}{2}$, and $\xi_\alpha = 1$ ($\alpha \simeq .16$) leads to $\beta = \frac{1}{4}$. The choice of δ has no influence on the leading term of the error. Numerical investigation, reinforced by some asymptotic considerations, indicates that it is useful to combine $\beta = \frac{1}{2}$ with $\delta = -1$ for accuracy in the tails, and $\beta = \frac{1}{4}$ with $\delta = 0$ for accuracy in the middle part of the distribution. Due to the influence of the term of order τ^{-2}, this "middle part" contains roughly the probabilities between .05 and .93, which is an asymmetric region. For something like $N > 1000$ the boundaries become .06 and .94.

For symmetric cases ($r = s$; for small N we might say $r \simeq s$), the asymptotic error depends in a somewhat intricate way on ξ and $\tilde{m}$ (or $\tilde{n}$). Now χ_t coincides with χ, and u_t with u, because $r = s = \frac{1}{2}N$ implies that $\tilde{n}+\tilde{r}-2\tilde{n}\tilde{r}=\frac{1}{2}$. In the doubly symmetric case $m = n = r = s$, one obtains that (2.27) with $\delta = 0$ coincides also with χ, regardless of the value of β. We have already observed that for symmetric tables χ is nearly always more accurate than u_h. Just as in the skew case, χ is asymptotically optimal at $\xi = \pm1$. However, (2.27) with suitable β and δ is expected to be still more accurate, as its coefficient of $\tau^{-2}\xi^3$ is smaller for any value of $\tilde{m}\tilde{n}$ (see Table 2.1). For asymptotic optimality at $\xi = \pm\xi_\alpha$ one may take $\beta = (\xi_\alpha^2+2)/24$ and $\delta = (1-\xi_\alpha^2)/6$. This means $\beta = 1/8$, $\delta = 0$ for $\xi_\alpha = 1$, and $\beta = \frac{1}{4}$, $\delta = -\frac{1}{2}$ for $\xi_\alpha = 2$. However, one may continue to use (2.27) with $\beta = \frac{1}{2}$, $\delta = -1$ or $\beta = \frac{1}{4}$, $\delta = 0$, just as in the skew case, without much loss of accuracy. The symmetry $r = s$ makes the error $O(\tau^{-2})$ for all ξ, and it may not be worthwhile to reduce the error to $o(\tau^{-2})$ for some special values of ξ. Besides, asymptotic optimality at $\xi = \pm2$ is also reached for $\beta = \frac{1}{2}$, $\delta = 1.5 - 1/(2\tilde{m}\tilde{n})$, and such a δ lies not too far from $\delta = -1$ unless $\tilde{n}$ is smaller than, say, .2.

Thus our main conclusion, supported by numerical evidence, is the advice to use the rather simple formula

$$(2.37) \qquad \Phi(2\{N-1\}^{-\frac{1}{2}}\{(k+1)^{\frac{1}{2}}(N-n-r+k+1)^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(r-k)^{\frac{1}{2}}\})$$

as a normal approximation to the hypergeometric distribution function $H_{n,r,N}(k)$. It is especially accurate near the customary significance levels; when the middle part of the distribution (say between .05 and .93) is more important, one may use

$$(2.38) \qquad \Phi(2N^{-\frac{1}{2}}\{(k+\tfrac{3}{4})^{\frac{1}{2}}(N-n-r+k+\tfrac{3}{4})^{\frac{1}{2}} - (n-k-\tfrac{1}{4})^{\frac{1}{2}}(r-k-\tfrac{1}{4})^{\frac{1}{2}}\}).$$

Although there are quite a few exceptions, we may safely say that (2.37) and (2.38) are generally superior to the classical normal and χ^2 approximations and to their modified forms χ_t and u_t: they are frequently better, sometimes much better, and rarely much worse. Moreover, their calculation is very easy. In the summary preceding Chapter I, we have already stated that the square root approximations to binomial and Poisson distribution functions can be obtained from (2.37) and (2.38) by obvious limiting processes.

2e. BETTER NORMAL APPROXIMATIONS

Improved normal approximations can be obtained from the simple ones by the addition of correction terms, or by the use of continuity corrections depending on some simple deviate. In this subsection both methods will be discussed. However, it should be emphasized that substantial improvement is difficult in the situations where it is most urgently needed. Indeed when simple normal approximations are grossly inaccurate, the parameter τ is usually so small that the error is not adequately described by the first term of its series expansion, certainly not for large values of $|\xi|$. In such a case removal of the leading term by a correction may not lead to a substantial gain in accuracy.

Table 2.2 has been included in order to illustrate that τ is small for many 2×2 tables. It gives also the value of NICHOLSON's parameter τ^* discussed on the next page.

TABLE 2.2

Values of $\tau = (mnrsN^{-3})^{\frac{1}{2}}$ and
$\tau^{*}=\{(m+1)(n+1)(r+1)(s+1)(N+2)^{-3}\}^{\frac{1}{2}}$
for some parameter triplets n,r,N

n	r	N	τ	τ^{*}
2	2	20	.40	.55
2	10	20	.67	.80
10	10	20	1.12	1.17
4	4	200	.28	.34
4	50	200	.86	.96
20	20	200	1.27	1.32
20	100	200	2.12	2.17
50	50	200	2.65	2.68
50	100	200	3.06	3.09
100	100	200	3.54	3.55

As far as we know, the only publication on improved normal approximations to the hypergeometric distribution function is NICHOLSON (1956); it is an extension of FELLER's (1945) results for the binomial case. NICHOLSON's main result can be described in the following notation:

$$\begin{aligned}
&\tilde{n}^{*} = (n+1)/(N+2);\ \tilde{r}^{*} = (r+1)/(N+2);\\
&\mu^{*} = (n+1)\tilde{r}^{*};\ \tau^{*2} = (1-\tilde{n}^{*})(1-\tilde{r}^{*})\mu^{*};\\
(2.39)\qquad &u_i = (i-\mu^{*})/\tau^{*};\\
&v_i = u_i - \tau^{*-1}(2\tilde{n}^{*}-1)(2\tilde{r}^{*}-1)(u_i^2+2)/6;\\
&R = 5\tau^{*-2}(1-\tilde{n}^{*}+\tilde{n}^{*2})(1-\tilde{r}^{*}+\tilde{r}^{*2})/36 + 2/(3N+6).
\end{aligned}$$

If now

$$(2.40)\qquad \tau^{*} > 3,\ j \geq \mu^{*},\ k \leq \mu^{*} - \tfrac{1}{2} + 2\tau^{*2}/3,\ k \leq n-4,$$

then

$$(2.41) \qquad P[\underline{j} \le \underline{a} \le \underline{k}] \approx \frac{N+1}{N+2}\exp(R)\,\{\Phi(v_{k+1}) - \Phi(v_j)\}.$$

Upper and lower bounds for $P[j \le \underline{a} \le k]$ are obtained by replacing v_j and v_{k+1} in (2.41) by slightly different expressions.

As can be seen from Table 2.2, the condition $\tau^* > 3$ is somewhat restrictive: it means e.g. $N > 142$ for $n = r = \frac{1}{2}N$ and $N > 254$ for $n = r = \frac{1}{4}N$. The conditions on j and k are still more restrictive. We have the impression that NICHOLSON was mainly interested in the upper and lower bounds mentioned above. For an approximation to the distribution function $H_{n,r,N}(k)$, we have found that

$$(2.42) \qquad \Phi\left(\frac{N+1}{N+2}\exp(R)\; v_{k+1}\right)$$

is rather more accurate than

$$(2.43) \qquad \frac{N+1}{N+2}\exp(R)\;\Phi(v_{k+1}).$$

Our numerical investigations show that (2.42) is also superior to (2.43) for approximations of $P[j \le \underline{a} \le k]$ in situations where (2.40) is fulfilled. This can be explained from the asymptotic expansions; straightforward but tedious calculations show that

$$(2.44) \qquad \begin{aligned} u_{k+1} &= \xi + \tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})(\xi^2 + 2)/6 + \\ &+ \tau^{-2}\,\{\xi^3(-1-2\widetilde{mn}-2\widetilde{rs}+26\widetilde{mnrs}) + \xi(-2-22\widetilde{mn}-22\widetilde{rs}+142\widetilde{mnrs})\}/72 + \\ &+ o(\tau^{-2}); \\ v_{k+1} &= \xi + \tau^{-2}\,\{\xi^3(-5+14\widetilde{mn}+14\widetilde{rs}-38\widetilde{mnrs}) + \\ &+ \xi(-10+10\widetilde{mn}+10\widetilde{rs}+14\widetilde{mnrs})\}/72 + o(\tau^{-2}); \\ \frac{N+1}{N+2}\exp(R) &= 1 + \tau^{-2}(10-10\widetilde{mn}-10\widetilde{rs}-14\widetilde{mnrs})/72; \\ (2.42) &= \Phi(\xi) + \tau^{-2}\,\phi(\xi)\,\xi^3(-5+14\widetilde{mn}+14\widetilde{rs}-38\widetilde{mnrs})/72 + o(\tau^{-2}); \\ (2.43) &= \Phi(\xi) + \tau^{-2}\,\{\phi(\xi)\,\xi^3(-5+14\widetilde{mn}+14\widetilde{rs}-38\widetilde{mnrs}) + \\ &+ (\Phi(\xi) - \xi\,\phi(\xi))(10-10\widetilde{mn}-10\widetilde{rs}-14\widetilde{mnrs})\}/72 + o(\tau^{-2}). \end{aligned}$$

For most choices of ξ and of the parameters n,r,N, (2.43) will have a much larger error. However, even (2.42) is by no means the best improved normal approximation. In a detailed investigation, we have examined several normal deviates with correction terms or variable continuity corrections. It turns out that the reasonably simple deviate

$$\chi + (\chi^2-1)\{-\tau^{-1}(\tilde{m}-\tilde{n})(\tilde{s}-\tilde{r})/6 + \tau^{-2}\chi(1-3\tilde{m}\tilde{n})/48\},$$
$$\text{where } \chi = (k+\tfrac{1}{2}-nrN^{-1})/\tau,\ \tau^2 = mnrsN^{-3},\ \tilde{m} = mN^{-1} \text{ etc.,} \tag{2.45}$$

is rather accurate, with some exceptions for small tails in skew tables. For rather skew tables, say $\tilde{n} \leq \tilde{r} \leq \frac{1}{4}$, it is usually better to replace (2.45) by (2.27) with $\beta = (\chi^2+2)/12$ and $\delta = 1.5 - 1/(2\tilde{m}\tilde{n})$.

Numerical information on errors of normal approximations to the hypergeometric distribution is given in section 5 of this Chapter.

3. POISSON APPROXIMATIONS

If $N \to \infty$ in such a way that $\mu = nrN^{-1}$ and $\tau^2 = mnrsN^{-3}$ have the same limit $\mu_0 (0 < \mu_0 < \infty)$, then (and only then) the distribution of the hypergeometric random variable $\underline{a}$ tends to the Poisson distribution with expectation μ_0 (VAN EEDEN & RUNNENBURG, 1960). Under these conditions $msN^{-2} \to 1$; as $m = N-n < N$ and $s = N-r < N$, it follows that $\tilde{m} = mN^{-1} \to 1$ and $\tilde{s} = sN^{-1} \to 1$, and thus $\tilde{n} = nN^{-1} \to 0$ and $\tilde{r} = rN^{-1} \to 0$. Because of our convention $n \leq r \leq \frac{1}{2}N$, one obtains from $\tilde{r} \to 0$ and $\mu = n\tilde{r} \to \mu_0 > 0$ that the smallest marginal n tends to infinity, and a fortiori $r \to \infty$, $s \to \infty$ and $m \to \infty$.

We are only interested in probabilities bounded away from 0 and 1 (cf. section I.3). Probabilities close to 0 cannot occur, as $P[\underline{a} = 0] \to e^{-\mu_0}$. In order to exclude probabilities close to 1, we shall restrict our considerations of the hypergeometric distribution function $H_{n,r,N}(k)$ to values of k not exceeding some positive constant K_0, say $K_0 = 4\mu_0 + 4\mu_0^{\frac{1}{2}}$. Since k and μ are now bounded, the O-symbols occurring in this section can be considered as uniform in k and μ, at least for fixed μ_0.

Our conditions include any situation for which $n \to \infty$ and $r \to \infty$ in such a way that $n \leq r$, $nrN^{-1} \to \mu_0$, $rN^{-1} \to 0$. Poisson approximations will be most suitable when e.g. $n = r = (\mu_0 N)^{\frac{1}{2}}$, but one also has a Poisson limit for $n = N^{\frac{1}{4}}$, $r = \mu_0 N^{\frac{3}{4}}$, or even $n = \log N$, $r = \mu_0 N/\log N$. However, for n bounded and $nrN^{-1} \to \mu_0$, the limit distribution is binomial (see section 4). As we do not know whether r is asymptotically of a larger order than n, we shall join terms of order n^{-1} and r^{-1} in our asymptotic expansions. When the remainder term is written as $O(n^{-3})$ this is sufficient to cover also terms proportional to r^{-3}, $n^{-1}r^{-2}$ or $n^{-2}r^{-1}$, because of our convention $n \leq r$. As $nrN^{-1} \to \mu_0$ with $0 < \mu_0 < \infty$, terms proportional to nN^{-1} or rN^{-1} are $O(n^{-1})$.

We shall now derive an asymptotic expansion for the exact Poisson parameter $\lambda_0 = \lambda_0(k,n,r,N)$ defined by $F_{\lambda_0}(k) = H_{n,r,N}(k)$; we recall that F_λ denotes the Poisson distribution function with parameter λ. The derivation will be followed by a search for simple functions of k,n,r,N which agree with λ_0 as far as possible. The section ends with the results of a numerical investigation of the accuracy of Poisson approximations with such simple functions as parameter. With a parameter depending also on k, they are far more accurate than the classical Poisson approximation with parameter nrN^{-1}. However, even such improved Poisson approximations are nearly always inferior to the binomial approximations discussed in section 4.

THEOREM

The exact Poisson parameter λ_0 defined by $F_{\lambda_0}(k) = H_{n,r,N}(k)$ satisfies the expansion

$$
\begin{aligned}
\lambda_0 = {} & \mu + (n^{-1}+r^{-1})\mu(\mu-k)/2 + \\
& + (n^{-2}+r^{-2})\mu(8\mu^2-7\mu k-k^2-2k)/24 + \\
& + n^{-1}r^{-1}\mu(2\mu^2-3\mu k+k^2-2\mu)/4 + \\
& + O(n^{-3}),
\end{aligned} \tag{3.1}
$$

where $\mu = nrN^{-1}$.

PROOF. For any j satisfying $0 \leq j \leq k \leq K_0$ we write

$$\log P[\underline{a} = j] = \log \{\binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n}\} = -\log j! + \tag{3.2}$$

$$- \log \frac{N!}{(N-r)!} + \log \frac{(N-n)!}{(N-n-r+j)!} + \log \frac{n!}{(n-j)!} + \log \frac{r!}{(r-j)!}.$$

By expansion of $\log (1-iN^{-1})$ and calculation of $\sum i$, $\sum i^2$ and $\sum i^3$, one obtains that

$$- \log \frac{N!}{(N-r)!} = -r \log N - \sum_{i=1}^{r-1} \log (1 - \frac{i}{N}) = \tag{3.3}$$

$$= - r \log N + \frac{r(r-1)}{2N} + \frac{r(r-\frac{1}{2})(r-1)}{6N^2} + \frac{r^2(r-1)^2}{12N^3} + O(n^{-3}).$$

After a similar evaluation of the last three terms in (3.2), some calculation leads to

$$P[\underline{a} = j] = \frac{e^{-\mu}\mu^j}{j!} [1 + \frac{n+r}{N}\{j - \tfrac{1}{2}\mu - \frac{j(j-1)}{\mu}\} + \tag{3.4}$$

$$+ \frac{1}{N}\{- \frac{j(j+1)}{2} + (\frac{n}{r} + 2 + \frac{r}{n})\frac{j\mu}{2} + \frac{\mu}{2} - (\frac{n}{r} + \frac{r}{n})\frac{(2j-1)j(j-1)}{12\mu}\} +$$

$$+ \tfrac{1}{2}(\frac{n+r}{N})^2\{j - \tfrac{1}{2}\mu - \frac{j(j-1)}{\mu}\}^2 - \frac{\mu}{6N^2}(2n^2+3nr+2r^2) + O(n^{-3})].$$

Next we sum (3.4) for $0 \leq j \leq k$, a finite number of terms, and use relations like

$$\sum_{j=0}^{k} j(j-1)\mu^j/j! = \mu^2 \sum_{i=0}^{k-2} \mu^i/i! \tag{3.5}$$

The final result is

$$\sum_{j=0}^{k} P[\underline{a} = j] = \sum_{j=0}^{k} e^{-\mu}\mu^j/j! + \tag{3.6}$$

$$+ \frac{e^{-\mu}\mu^k}{k!} [\frac{(n+r)(k-\mu)}{2N} + (n^{-2}+r^{-2})\{- k(k-1)(k-2)\frac{\mu}{8} +$$

$$+ k(k-1)(\frac{3\mu^2}{8} - \frac{\mu}{3}) + k(\frac{2\mu^2}{3} - \frac{3\mu^3}{8}) - \frac{\mu^3}{3} + \frac{\mu^4}{8}\} +$$

$$+ n^{-1}r^{-1}\{- k(k-1)(k-2) \frac{\mu}{4} + k(k-1)(\frac{3\mu^2}{4} - \mu) + k(- \frac{3\mu^3}{4} + \frac{3\mu^2}{2} - \frac{\mu}{2}) +$$

$$+ \frac{\mu^4}{4} - \frac{\mu^3}{2} + \frac{\mu^2}{2}\}] + O(n^{-3}).$$

The definition of the exact Poisson parameter λ_0 means that (3.6) must be equal to $F_{\lambda_0}(k)$. Let us suppose that $\lambda_0 = \mu+\varepsilon+\delta$, where ε is asymptotically of the order n^{-1} and δ is asymptotically of the order n^{-2}. Then, because of II (1.4),

$$
\begin{aligned}
F_{\lambda_0}(k) &= \sum_{j=0}^{k} e^{-\lambda_0} \lambda_0^j/j! = \int_{\lambda_0}^{\infty} e^{-t}t^k/k!\, dt = \\
&= \int_{\mu}^{\infty} e^{-t}t^k/k!\, dt - \int_{\mu}^{\mu+\varepsilon+\delta} e^{-t}t^k/k!\, dt = \\
&= \sum_{j=0}^{k} \frac{e^{-\mu}\mu^j}{j!} - (\varepsilon+\delta)\frac{e^{-\mu}\mu^k}{k!} - \frac{\varepsilon^2(k-\mu)e^{-\mu}\mu^{k-1}}{2k!} + O(n^{-3}).
\end{aligned}
\tag{3.7}
$$

Combining (3.6) and (3.7) one obtains

$$
\begin{aligned}
\varepsilon &= N^{-1}(n+r)(\mu-k)/2, \\
\delta &= \mu^{-1}N^{-2}(n+r)^2(8\mu^2-7\mu k-k^2-2k)/24 + \\
&\quad + N^{-1}(-\mu^2-\mu k+2k^2-3\mu+k)/6 + O(n^{-3}),
\end{aligned}
\tag{3.8}
$$

which leads directly to a proof of (3.1).

Now we want to find a simple function of n, r, N and k that agrees as far as possible with the expansion (3.1) of the exact Poisson parameter λ_0. We may replace the classical $\mu = nrN^{-1}$ by $\lambda_a = \{2nr + (n+r)(\mu-k)\}/\{2N\}$; it is clear that this reduces the error from $O(n^{-1})$ to $O(n^{-2})$. Still considerably better, and also rather simple, is

$$
\lambda_b = \frac{(2n-k)(2r-k)}{2(2N-n-r+1)} = \frac{(2n-k)(2r-k)}{2(m+s+1)}\,, \tag{3.9}
$$

for which

$$
\lambda_b - \lambda_0 = (n^{-2}+r^{-2})\mu(-2\mu^2+k\mu+k^2+2k)/24 + n^{-1}r^{-1}\mu^2k/4 + O(n^{-3}). \tag{3.10}
$$

A numerical investigation confirms that λ_b is generally superior to λ_a, which is in turn superior to μ. The computer results indicate that λ_b gives much better results for small left hand tails ($k < \mu$) than for small right hand tails ($k > \mu$). Inserting $k = \mu \pm d$ into (3.10), one finds that this behaviour can indeed be predicted from the leading term of the error.

The numerical results demonstrate that all Poisson approximations become rapidly worse for increasing values of the marginal fractions $\tilde{n} = nN^{-1}$ and $\tilde{r} = rN^{-1}$. Even for λ_b, the errors can be considered to be unacceptably large as soon as $\tilde{r} > .1$ (we recall that $\tilde{n} \leq \tilde{r} \leq \frac{1}{2}$ because of our convention $n \leq r \leq \frac{1}{2}N$). For fixed values of $\tilde{n}$ and $\tilde{r}$, the value of the grand total N has little influence on the accuracy of Poisson approximations.

Even when λ_b differs little from the exact parameter λ_0, the difference between $F_{\lambda_b}(k)$ and $F_{\lambda_0}(k)$ can be rather large. Moreover, the convergence of the expansion (3.1) is not very fast. This makes Poisson approximations to the hypergeometric distribution function somewhat unattractive. When binomial tables are available, it is best to use binomial approximations discussed in the next section for skew tables or tables with a small grand total N. Otherwise one may use the normal square root approximations (2.37) and (2.38), or (2.45) for extra accuracy.

When binomial tables are not available but Poisson tables are, it would be attractive to use a Poisson approximation that is accurate for moderately skew tables. Consideration of graphs displaying the exact Poisson parameter λ_0 as a function of k for various parameter triplets (n,r,N) has led us to consider

$$(3.11) \qquad \lambda_c = \mu + (\mu-k)(2r-n+10\mu)/(3N).$$

For this empirically determined parameter, one finds from (3.1) that

$$(3.12) \qquad \lambda_c - \lambda_0 = \left(\frac{1}{n} - \frac{5}{r}\right)\mu(\mu-k)/6 + O(n^{-2}).$$

For small $\tilde{n}$ and $\tilde{r}$, the numerical investigation shows that λ_c is less accurate than λ_b. As the error of λ_c is $O(n^{-1})$ $\left[\text{although } O(n^{-2}) \text{ for } n = \frac{r}{5}\right]$ against $O(n^{-2})$ for λ_b, this behaviour could be more or less expected. Unless N is very large, there will be a region of moderately skew tables for which $F_{\lambda_c}(k)$ is superior not only to $F_{\lambda_b}(k)$, but also to the square root normal approximations (2.37) and (2.38). Some examples of parameter triplets belonging to

TABLE 3.1. Examples of parameter triplets (n, r, N) for which the approximation $F_{\lambda_c}(k)$ is superior to $F_{\lambda_b}(k)$, but also to (2.37) and (2.38)

n	r	N	n	r	N
8	20	80	4	80	200
8	32	80	20	50	200
8	40	80	20	80	200
4	20	200	10	200	800
4	50	200	80	200	800

this region are given in Table 3.1. We repeat, however, that all Poisson approximations are generally inferior to the binomial approximations $G_{n,w}(k)$, and even more $G_{n,w+\gamma}(k)$, discussed in the next section.

4. BINOMIAL APPROXIMATIONS

A binomial approximation to the hypergeometric distribution has a somewhat limited applicability: it can only be used when binomial tables are available and contain the necessary values. Our experience indicates that it is seldom advisable to use a binomial approximation when the binomial probability itself must be obtained from a normal or Poisson approximation. In such cases, a direct normal or Poisson approximation to the hypergeometric probability is nearly always more successful. Moreover, binomial approximations will be often used when n is small and p is far from 0 or 1, and such parameter values are unfavourable for Poisson or normal approximations to the binomial.

In this section it is assumed that $N \to \infty$ in such a way that $\mu = nrN^{-1} \to$ $\to \mu_0$ and $\tau^2 = mnrsN^{-3} \to \tau_0^2$, where $0 < \tau_0^2 < \mu_0 < \infty$. It is easily seen that this implies that $\tilde{n}\tilde{r} = \mu N^{-1} \to 0$ (we recall that $\tilde{n} = nN^{-1}$ and $\tilde{r} = rN^{-1}$). As we have ordered our marginals such that $n \leq r \leq \frac{1}{2}N$, it follows that $\tilde{n} \to 0$ and $\tilde{m} = 1-\tilde{n} \to 1$. Thus $\tilde{s} = \tau^2/(\tilde{m}\mu) \to \tau_0^2/\mu_0$, which implies $\tilde{r} = 1-\tilde{s} \to 1 - \tau_0^2/\mu_0$ and $n = \mu/\tilde{r} \to \mu_0^2/(\mu_0-\tau_0^2)$. Thus our assumptions imply that the smallest marginal n tends to a finite positive limit. This limit is then automatically an integer, n_0 say. We have $n = n_0$ for sufficiently large N, and the fraction $\tilde{r}$ tends to the constant $\mu_0/n_0 = 1 - \tau_0^2/\mu_0$. As $k \leq n$, we obtain that k is bounded. O-symbols in this section can be considered as uniform in k and n.

Since many years, the binomial distribution with parameters n and $\tilde{r}$ is used as an approximation to the hypergeometric (n,r,N) distribution. We recall that $G_{n,p}(k)$ denotes the binomial distribution function, see III (1.1). The approximation is nearly always conservative (it overestimates both hypergeometric tails). Its use means neglecting the fact that sampling is without replacement. Some authors say that it can safely be used when the sampling fraction $\tilde{n}$ does not exceed .1. The example $H_{20,80,200}(3) = .0121$, $G_{20,.4}(3) = .0160$ may serve to indicate that the classical binomial approximation grossly overestimates hypergeometric tails for $\tilde{n} = .1$ (more examples can be found in section 5).

Let us now review some improved binomial approximations. Expressing the hypergeometric distribution function as a double contour integral, WISE (1954) obtains a quickly converging series expansion in terms of incomplete beta functions. We have rewritten it in terms of binomial distribution functions, cf. III (1.2), and explicitly calculated one extra term which follows easily from WISE's results. In the notation

$$M = 2N-n+1, \qquad w = (2r-k)/M \tag{4.1}$$

one obtains from WISE (1954) that

$$\begin{aligned} H_{n,r,N}(k) = &\Big[1 + M^{-2}(n+1)n(n-1)/6 + \\ &+ M^{-4}(5n-7)(n+3)(n+2)(n+1)n(n-1)/90 + O(M^{-6})\Big] \times \\ &\times\Big[G_{n,w}(k) + M^{-2}n(n-1)\{(-n+k-1)G_{n-2,w}(k) + \\ &+ 2G_{n-2,w}(k-1) - (k+2)G_{n-2,w}(k-2)\}/6 + \\ &+ M^{-4}n(n-1)(n-2)(n-3)\{(5(n-k)^2+2(n-k)+7)G_{n-4,w}(k) + \\ &- 28G_{n-4,w}(k-1) + 42G_{n-4,w}(k-2) - 28G_{n-4,w}(k-3) + \\ &+ (5k^2+12k+8)G_{n-4,w}(k-4)\}/360 + O(M^{-6})\Big]. \end{aligned} \tag{4.2}$$

WISE observes that the new parameter $w = (r-\frac{1}{2}k)/(N-\frac{1}{2}n+1)$ can be considered, in the terminology of model (i) of section 1, as a kind of average fraction of red balls during an experiment consisting of n drawings without replacement resulting in k red balls. As an approximation to $H_{n,r,N}(k)$, WISE suggests to use $G_{n,w}(k)$ or

$$(4.3)\quad \begin{aligned} &G_{n,w}(k) + M^{-2}n(n-1)\{(n+1)G_{n,w}(k) - (n-k+1)G_{n-2,w}(k) + \\ &+ 2\, G_{n-2,w}(k-1) - (k+2)G_{n-2,w}(k-2)\}/6. \end{aligned}$$

Even for $\tilde{n}$ close to $\frac{1}{2}$, (4.3) is rather accurate. However, it is unattractive to evaluate four different values of the binomial distribution function, the more so because w will usually have values for which interpolation is required. By a Taylor expansion, we shall reduce (4.3) to one binomial distribution function with a success parameter $w+\delta$. One has

$$(4.4)\quad G_{n,w+\delta}(k) = G_{n,w}(k) + \delta\,\frac{\partial G_{n,w}(k)}{\partial w} + O(\delta^{-2}),$$

and

$$(4.5)\quad \frac{\partial G_{n,w}(k)}{\partial w} = -\frac{n!}{k!(n-k-1)!}\,w^k(1-w)^{n-k-1},$$

because the binomial distribution function equals the incomplete beta function ratio, cf. III (1.2). For a comparison of (4.3) and (4.4) we have to work out (4.3). The factor between curly brackets can be written as

$$(4.6)\quad \begin{aligned} &(n+1)\{G_{n,w}(k) - G_{n-2,w}(k-2)\} + \\ &- (n-k-1)P[\underline{y}_{n-2} = k-1] - (n-k+1)P[\underline{y}_{n-2} = k], \end{aligned}$$

where $\underline{y}_h$ denotes the number of successes in h independent experiments with success probability w. Now

$$P[\underline{y}_n \leq k | \underline{y}_{n-1} \leq k-1] \times P[\underline{y}_{n-1} \leq k-1] = 1 \times G_{n-1,w}(k-1);$$

$$(4.7) \qquad P[\underline{y}_n \leq k | \underline{y}_{n-1} = k] \times P[\underline{y}_{n-1} = k] = (1-w) \times \binom{n-1}{k} w^k (1-w)^{n-k-1};$$

$$P[\underline{y}_n \leq k | \underline{y}_{n-1} > k] \times P[\underline{y}_{n-1} > k] = 0.$$

Addition gives

$$(4.8) \qquad G_{n,w}(k) = P[\underline{y}_n \leq k] = G_{n-1,w}(k-1) + \binom{n-1}{k} w^k (1-w)^{n-k}.$$

By repeated application of (4.8) one reduces (4.6) to an expression consisting of individual binomial terms multiplied by simple factors. The result, similar to formula (4.11) of WISE (1954), can be compared to (4.4), and one obtains finally that

$$(4.9) \qquad \delta = M^{-2}[k(n+1)\{w - w^{-1}\} - (n-k-1)(n+1)\{1-w - (1-w)^{-1}\} + k(n-k-1)\{w^{-1} - (1-w)^{-1}\}]/6 + O(M^{-4}).$$

One may use $G_{n,w+\delta}(k)$, with δ given by (4.9) and M and w given by (4.1), as a binomial approximation to the hypergeometric distribution function $H_{n,r,N}(k)$. As δ is still somewhat unwieldy for hand calculation, we propose to replace it by

$$(4.10) \qquad \gamma = M^{-2}n(2\mu-2k-1)/3;$$

we recall that $\mu = nrN^{-1}$ and $M = 2N-n+1$. Numerical and theoretical investigations show that $w+\gamma$ and $w+\delta$ are usually almost equal. Thus $G_{n,w+\gamma}(k)$ is a relatively simple and rather accurate approximation; $G_{n,w}(k)$ will usually be accurate enough when the sampling fraction $\tilde{n}$ remains below .1. The numerical examples in section 5 illustrate the spectacular gain in accuracy compared to the classical binomial approximation $G_{n,\tilde{r}}(k)$.

SANDIFORD (1960) has proposed to use the binomial distribution for which the first two moments agree with the hypergeometric distribution. This means the use of $G_{n^*,p^*}(k)$, where

$$(4.11)\qquad \begin{aligned} n^{*} &= \text{nearest integer to } nr(N-1)/\{N(N-1)-ms\},\\ p^{*} &= \mu/n^{*}; \end{aligned}$$

we recall that $\mu = nrN^{-1}$, $m = N-n$ and $s = N-r$. One has $n^{*} \leq n$ and usually $n^{*} < n$; in certain cases it may be a drawback that the values exceeding n^{*} get probability zero according to the approximation. In an extensive numerical investigation, we have found that $G_{n^{*},p^{*}}(k)$ is nearly always less accurate than $G_{n,w}(k)$, which contradicts the opinion of HALD (1967) that WISE's approximation is generally inferior to SANDIFORD's.

BOLSHEV (1964a) obtains that

$$(4.12)\qquad H_{n,r,N}(k) \simeq I_{1-x}(n^{**}-k+c,k-c+1) = G_{n^{**},x}(k-c),$$

where

$$(4.13)\qquad \begin{aligned} n^{**} &= (N-2)^{2}mnrs/\left[(N-1)(ms+nr-N)\{N(n+r-1) - 2nr\}\right],\\ x &= \left[N(n+r-1) - 2nr\right]/\left[N(N-2)\right],\\ c &= n(n-1)r(r-1)/\left[(N-1)(ms+nr-N)\right]. \end{aligned}$$

Obviously n^{**} and c will usually not have integer values. This is no drawback if tables of the incomplete beta function I are used, but these tables are less widespread than binomial tables and if they are available, interpolation is cumbersome. In our numerical investigation, we have therefore rounded off n^{**} to the nearest integer. As rounding off $(k-c)$ could introduce rather large errors, we have not only considered a solution with $(k-c)$ rounded off to the nearest integer, but also a calculation with linear interpolation between the nearest integers above and below $(k-c)$. Just as for n^{*}, one usually has $n^{**} < n$, with the drawback of zero probability for values exceeding n^{**}.

ORD (1968) obtains that

$$(4.14)\qquad \begin{aligned} H_{n,r,N}(k) &= G_{n,\tilde{r}}(k) +\\ &- \frac{n(n-1)r}{2Ns}\{G_{n,\tilde{r}}(k) - 2\,G_{n-1,\tilde{r}}(k-1) + G_{n-2,\tilde{r}}(k-2)\} + O(N^{-2}). \end{aligned}$$

He considers also the case that r is bounded and $n \to \infty$. We have excluded this by our convention $n \le r \le \frac{1}{2}N$: in the terminology of model (i) of section 1, we may have to interchange "drawn balls" with "red balls", "red balls" with "black balls" or "drawn balls" with "not drawn balls" in order to comply with this convention. ORD's formula for r bounded and $n \to \infty$ coincides with (4.14) after interchange of n and r, at least after an obvious redefinition of his symbol B.

Just as for WISE's formula (4.3), we would like to avoid the evaluation of many values of the binomial distribution function in (4.14). The same Taylor series argument and use of (4.8) with w replaced by $\tilde{r}$ lead to the conclusion that

$$(4.15) \qquad G_{n,\tilde{r}+\beta}(k), \text{ with } \beta = N^{-1}\left[(n-1)\tilde{r} - k\right]/2,$$

agrees with (4.14) apart from a different $O(N^{-2})$ term. A little calculation shows that $\tilde{r}+\beta = w+O(N^{-2})$, i.e. the approximations of ORD and WISE agree up to terms of order N^{-2}.

Numerical investigation shows that ORD's approximation (4.14) is usually slightly better than WISE's $G_{n,w}(k)$, but inferior to $G_{n,w+\gamma}(k)$, and even more to $G_{n,w+\delta}(k)$ and (4.3). The modified form (4.15) of ORD's result is even less accurate than $G_{n,w}(k)$; the same holds for the proposals of SANDIFORD and BOLSHEV. The latter might be better when fractional arguments could be used.

Practically without exceptions, (4.3) is better than $G_{n,w+\delta}(k)$, but both are extremely accurate: for sampling fraction $\tilde{n} \le \frac{1}{4}$ and tails of at least .001, the relative tail error is nearly always less than 1 per cent. With a few exceptions, occurring for right hand tails between .001 and .008, any tail of at least .001 is given for $\tilde{n} \le .1$ with a relative error below $\frac{1}{2}$ per cent. by $G_{n,w+\delta}(k)$, below 1 per cent. by $G_{n,w+\gamma}(k)$ and below 2 per cent. by $G_{n,w}(k)$. The relatively simple $G_{n,w+\gamma}(k)$ remains accurate in the unfavourable situation $\tilde{n} = \frac{1}{2}$: relative tail errors exceeding 5 per cent. are then exceptional for tails of at least .001.

We suggest to use binomial approximation with the given value of n, and with success probability $p = w$ for $\tilde{n} \le .1$ and $p = w+\gamma$ otherwise, see

(4.1) and (4.10). For very accurate results one may use $p = w+\delta$. The choice of a binomial parameter n differing from the hypergeometric smallest marginal n (SANDIFORD, BOLSHEV) is not recommended.

5. GENERAL ADVICE AND NUMERICAL INFORMATION

Table 5.1 gives a general advice on normal, Poisson and binomial approximations to the hypergeometric distribution function. A simple recommendation inevitably contains some subjective elements; we refer to the summary and to Chapter I for a general outline of our view.

Next the relative tail error I (2.3) is tabled for some approximations selected from the previous sections. For fixed values of the marginal fractions $\tilde{n} = nN^{-1}$ and $\tilde{r} = rN^{-1}$ only the distributions with $N = 20$ and $N = 200$ are tabled: the presence of three parameters n, r, N makes it difficult to condense the numerical information into a few pages.

TABLE 5.1. Advice for approximation to the hypergeometric distribution function $H_{n,r,N}(k) = \sum_{j=0}^{k} \binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n}$. In most statistical applications, accurate approximation to probabilities between .005 and .05 or between .95 and .995 will be essential. In such cases, one may use the suggestion marked "for tails". Φ denotes the standard normal distribution function I (2.1).

When cumulative binomial tables can be applied, use

$$\sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}$$

with (4.1) $p = (2r-k)/(2N-n+1)$ for quick work,

(4.10) $p = \frac{2r-k}{2N-n+1} - \frac{2n(k+\frac{1}{2}-nrN^{-1})}{3(2N-n+1)^2}$ for accurate results.

Without binomial tables, use for quick work

(2.37) $\Phi(2\{N-1\}^{-\frac{1}{2}}\{(k+1)^{\frac{1}{2}}(N-n-r+k+1)^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(r-k)^{\frac{1}{2}}\})$ for tails,

(2.38) $\Phi(2N^{-\frac{1}{2}}\{(k+\frac{3}{4})^{\frac{1}{2}}(N-n-r+k+\frac{3}{4})^{\frac{1}{2}} - (n-k-\frac{1}{4})^{\frac{1}{2}}(r-k-\frac{1}{4})^{\frac{1}{2}}\})$ for values between .05 and .93,

and for accurate results (2.45), see p.136.

When cumulative Poisson tables are available but binomial tables are not, one may use

$$\sum_{j=0}^{k} e^{-\lambda}\lambda^j/j!$$

with (3.9) $\lambda = \frac{1}{2}(2n-k)(2r-k)/(2N-n-r+1)$ for $nN^{-1} \leq rN^{-1} \leq .1$,

(3.11) $\lambda = \mu + (\mu-k)(2r-n+10\mu)/(3N)$, where $\mu = nrN^{-1}$, otherwise.

We recall that the hypergeometric parameters are assumed to satisfy $n \leq r \leq \frac{1}{2}N$. *Rows and columns of a 2×2 table can always be rearranged so as to comply with this convention, see p. 116.*

TABLE 5.2. Event a ≤ k or a ≥ k+1, exact hypergeometric probability and relative tail error in per cent. for some normal, Poisson and binomial approximations. For distributions with N = 2r, probabilities and errors are the same for a ≤ k and a ≥ n-k, provided that Poisson approximations are applied to a ≤ n-k-1 when P[a ≤ k] must be evaluated for k ≥ n/2. Example : for $H_{4,100,200}(0)$ = .0606, (2.37) has a relative error of +23.80 per cent., it gives (1+.2380)×.0606 = .0750

Event	Probability	impr. (2.45)	square root (2.37)	square root (2.38)	χ (2.3)	u_h (2.1)	χ_t p.131	Poisson μ=nr/N	(3.11) λ_c	(3.9) λ_b	binomial $\hat{r}$=r/N	(4.1) w	(4.10) w+γ	(4.9) w+δ
n = 4 r = 4 N = 200														
a ≥ 1	.0782	+24.17	+2.78	+54.72	-17.07	-16.45	+614.18	-1.69	-.93	-.01	-.73	-.00	-.04	-.00
a ≥ 2	.0018	-100.00	+401.94	+649.36	-99.99	-99.99	-84.89	+70.02	+40.86	+4.46	+30.92	+2.08	+1.80	+.01
n = 4 r = 20 N = 200														
a ≥ 1	.3461	+2.91	-22.50	-1.63	+25.15	+25.20	+97.55	-4.75	+.35	-.08	-.64	-.00	-.00	-.00
a ≥ 2	.0506	+13.18	+7.35	+37.65	-36.68	-36.02	+125.47	+21.72	+1.17	+2.20	+3.43	+.06	+.03	+.00
a ≥ 3	.0032	-85.99	+129.45	+191.18	-93.73	-93.52	-40.50	+144.10	+4.13	+13.32	+13.94	+.27	+.16	+.00
n = 4 r = 50 N = 200														
a ≤ 0	.3132	-2.77	+23.69	+4.43	-10.64	-10.49	-38.05	+17.45	-1.57	+.82	+1.02	+.00	+.00	+.00
a ≥ 2	.2606	+.85	-9.58	+.60	+7.38	+7.57	+46.12	+1.38	+1.38	+1.18	+.41	+.01	-.00	+.00
a ≥ 3	.0490	+.21	+9.93	+19.35	-18.20	-17.43	+44.50	+63.84	+3.72	+11.11	+3.61	+.04	+.01	+.00
a ≥ 4	.0036	-28.15	+74.00	+67.69	-50.22	-49.04	+17.77	+433.33	+22.70	+60.68	+9.72	+.10	+.03	+.00
n = 4 r = 80 N = 200														
a ≤ 0	.1270	-1.56	+24.43	+3.05	+1.09	+1.56	-14.34	+58.98	+.50	+4.57	+2.05	+.01	-.00	+.00
a ≤ 1	.4743	-.33	+5.02	+1.12	-3.24	-3.22	-11.47	+10.67	-.65	+.40	+.18	+.00	+.00	+.00
a ≥ 3	.1769	-.35	+2.87	+3.16	-.08	+.26	+15.10	+22.49	+6.00	+7.40	+1.32	+.01	-.00	+.00
a ≥ 4	.0245	-2.70	+28.00	+11.18	+2.51	+3.69	+28.63	+222.34	+37.64	+52.04	+4.70	+.03	+.00	+.00
n = 4 r = 100 N = 200														
a ≤ 0	.0606	-.97	+23.80	+3.84	+6.99	+7.78	+6.99	+123.25	+8.67	+11.57	+3.10	+.02	-.00	+.00
a ≤ 1	.3106	-.39	+4.09	+1.52	-1.24	-1.10	-1.24	+30.72	+2.14	+3.21	+.61	+.00	-.00	+.00

TABLE 5.3. Event a ≤ k or a ≥ k+1, exact hypergeometric probability and relative tail error in per cent. for some normal, Poisson and binomial approximations. For distributions with N = 2r, probabilities and errors are the same for a ≤ k and a ≥ n-k, provided that Poisson approximations are applied to a ≤ n-k-1 when P[a ≤ k] must be evaluated for k ≥ n/2

Event	Probability	impr. (2.45)	square root (2.37)	square root (2.38)	χ (2.3)	u_h (2.1)	χ_t p.131	Poisson μ=nr/N	(3.11) λ_c	(3.9) λ_b	binomial $\tilde{r}$=r/N	(4.1) w	(4.10) w+γ	(4.9) w+δ
n = 2 r = 2 N = 20														
a ≥ 1	.1947	+4.09	-15.27	+8.42	+17.10	+20.04	+166.93	-6.92	-1.35	-.15	-2.43	-.07	-.31	-.00
a ≥ 2	.0053	-96.86	+106.92	+194.50	-88.23	-84.39	+41.54	+232.94	+85.43	+29.63	+90.00	+12.43	+9.12	+.35
n = 20 r = 20 N = 200														
a ≤ 0	.1085	+3.28	+38.16	-1.01	+9.91	+10.45	-29.64	+24.68	+9.12	+.46	+12.01	+.23	-.08	+.00
a ≤ 1	.3782	-.87	+17.30	+3.83	-8.19	-8.10	-31.33	+7.35	+2.66	-.08	+3.58	+.06	-.01	+.00
a ≥ 3	.3213	+.38	-12.90	-1.96	+8.06	+8.17	+38.11	+.62	+.62	+.62	+.54	+.08	-.00	+.00
a ≥ 4	.1222	+2.28	-9.17	+4.75	-2.41	-1.93	+44.74	+16.88	+7.21	+1.96	+8.76	+.43	+.04	+.00
a ≥ 5	.0345	+9.32	+3.17	+21.39	-28.32	-27.50	+25.60	+52.46	+19.94	+4.37	+25.01	+1.16	+.23	+.01
a ≥ 6	.0073	+21.69	+28.99	+54.41	-59.29	-58.42	-14.81	+126.23	+41.75	+8.28	+53.70	+2.42	+.70	+.02
a ≥ 7	.0012	+33.86	+79.03	+117.66	-82.65	-82.06	-56.41	+286.31	+77.91	+14.26	+103.31	+4.43	+1.60	+.04
n = 2 r = 5 N = 20														
a ≥ 1	.4474	+1.55	-11.67	-1.28	+11.76	+11.76	+41.87	-12.05	+1.87	-.60	-2.21	-.07	-.07	-.00
a ≥ 2	.0526	-5.68	+5.08	+21.58	-19.07	-11.27	+60.07	+71.39	+12.73	+15.45	+18.75	+1.18	+.42	+.00
n = 20 r = 50 N = 200														
a ≤ 0	.0023	+60.59	-6.08	-46.12	+217.83	+223.29	+133.65	+199.40	+1.33	+5.58	+40.91	+.77	-.35	+.00
a ≤ 1	.0194	+10.80	+11.02	-14.27	+46.06	+47.66	+13.25	+108.07	+.09	+3.07	+25.13	+.52	-.20	+.00
a ≤ 2	.0800	+1.50	+13.58	-1.87	+8.46	+9.14	-11.49	+55.79	-.60	+1.14	+14.06	+.31	-.10	+.00
a ≤ 3	.2112	-.13	+10.88	+1.66	-1.94	-1.66	-16.01	+25.48	-.87	-.21	+6.61	+.15	-.04	+.00
a ≤ 4	.4068	-.15	+7.02	+1.68	-3.45	-3.38	-13.56	+8.29	-.84	-1.00	+1.99	+.04	-.01	+.00
a ≥ 6	.3800	+.08	-6.05	-1.23	+3.35	+3.42	+14.51	+1.06	+1.06	+2.11	+.74	+.04	-.00	+.00
a ≥ 7	.2036	+.20	-5.86	+.10	+1.72	+2.00	+17.67	+16.80	+1.60	+4.66	+5.21	+.19	-.02	+.00
a ≥ 8	.0903	+.66	-3.54	+3.61	-3.91	-3.31	+16.58	+47.67	+2.06	+8.74	+12.73	+.44	-.02	+.00
a ≥ 9	.0329	+1.53	+1.67	+10.18	-13.86	-12.92	+9.94	+106.67	+2.31	+14.90	+24.21	+.81	-.00	+.00
a ≥ 10	.0098	+2.52	+ 10.80	+21.00	-27.27	-26.02	-2.13	+223.62	+2.17	+23.99	+40.97	+1.32	+.05	+.00
a ≥ 11	.0024	+2.92	+25.45	+37.87	-42.36	-40.92	-18.09	+473.12	+1.50	+37.29	+64.97	+2.00	+.15	+.01

TABLE 5.4. Event $\underline{a} \le k$ or $\underline{a} \ge k+1$, exact hypergeometric probability and relative tail error in per cent. for some normal, Poisson, and binomial approximations. For distributions with N = 2r, probabilities and errors are the same for $\underline{a} \le k$ and $\underline{a} \ge n-k$, provided that Poisson approximations are applied to $\underline{a} \le n-k-1$ when $P[\underline{a} \le k]$ must be evaluated for $k \ge n/2$

Event	Probability	impr. (2.45)	square root (2.37)	square root (2.38)	χ (2.3)	u_h (2.1)	χ_t p.131	Poisson $\mu=nr/N$	(3.11) λ_c	(3.9) λ_b	binomial $\tilde{r}=r/N$	(4.1) w	(4.10) $w+\gamma$	(4.9) $w+\delta$
n = 2 r = 8 N = 20														
$\underline{a} \le 0$	.3474	-2.03	+8.51	+1.63	-6.72	-5.52	-18.94	+29.35	-3.53	+2.54	+3.64	+.12	+.03	+.00
$\underline{a} \ge 2$	.1474	-2.61	+2.19	+4.13	-2.67	+1.53	+17.23	+29.75	+12.05	+11.84	+8.57	+.38	+.06	+.00
n = 20 r = 80 N = 200														
$\underline{a} \le 2$	.0024	+11.88	-1.47	-17.45	+67.10	+70.40	+49.02	+464.67	-3.52	+16.20	+48.27	+1.05	-.30	+.00
$\underline{a} \le 3$	.0121	+3.50	+4.22	-6.90	+25.49	+27.21	+13.81	+250.08	-3.57	+10.19	+31.85	+.74	-.20	+.00
$\underline{a} \le 4$	.0425	+.78	+5.80	-1.74	+8.38	+9.34	-.09	+134.24	-3.33	+5.40	+19.79	+.48	-.13	+.00
$\underline{a} \le 5$	.1130	+.04	+5.32	+.35	+1.34	+1.85	-5.10	+69.22	-2.88	+1.75	+11.14	+.28	-.07	+.00
$\underline{a} \le 6$	.2377	-.07	+4.02	+.85	-1.02	-.79	-5.93	+31.85	-2.32	-.82	+5.19	+.13	-.03	+.00
$\underline{a} \le 7$	.4101	-.04	+2.57	+.65	-1.26	-1.20	-4.88	+10.45	-1.72	-2.38	+1.41	+.03	-.01	+.00
$\underline{a} \ge 9$	.4005	+.02	-2.03	-.40	+1.11	+1.17	+4.85	+1.74	+1.74	+4.52	+.98	+.04	-.01	+.00
$\underline{a} \ge 10$	.2337	+.02	-1.73	+.10	+.67	+.91	+5.80	+21.27	+2.30	+9.54	+4.70	+.15	-.02	+.00
$\underline{a} \ge 11$	.1152	+.02	-.66	+1.26	-.62	-.12	+5.99	+59.76	+2.85	+17.69	+10.65	+.32	-.05	+.00
$\underline{a} \ge 12$	.0473	-.02	+1.38	+3.21	-2.64	-1.77	+5.48	+136.40	+3.41	+30.51	+19.39	+.56	-.07	+.00
$\underline{a} \ge 13$	.0160	-.19	+4.57	+6.07	-4.89	-3.58	+4.73	+299.42	+4.12	+50.64	+31.66	+.88	-.10	+.00
$\underline{a} \ge 14$	.0044	-.53	+9.06	+9.86	-6.53	-4.68	+4.67	+684.97	+5.26	+83.03	+48.49	+1.28	-.13	+.00
n = 2 r = 10 N = 20														
$\underline{a} \le 0$	.2368	-2.29	+5.43	+1.94	-3.72	-1.30	-3.72	+55.33	-2.60	+6.30	+5.56	+.21	+.03	+.00
n = 20 r = 100 N = 200														
$\underline{a} \le 4$	.0040	+1.75	+4.54	-1.68	+17.85	+20.10	+17.85	+624.14	+6.87	+27.25	+46.27	+1.12	-.21	+.00
$\underline{a} \le 5$	.0159	+.38	+4.21	+.32	+6.86	+8.27	+6.86	+322.99	+3.93	+17.15	+30.48	+.78	-.15	+.00
$\underline{a} \le 6$	.0485	+.00	+3.17	+.87	+2.01	+2.89	+2.01	+168.31	+1.71	+9.25	+18.88	+.50	-.10	+.00
$\underline{a} \le 7$	.1190	-.05	+2.00	+.76	+.21	+.71	+.21	+84.99	+.12	+3.30	+10.54	+.29	-.06	+.00
$\underline{a} \le 8$	.2402	-.03	+.98	+.42	-.18	+.05	-.18	+38.57	-.93	-.89	+4.80	+.13	-.03	+.00
$\underline{a} \le 9$	.4071	-.01	+.25	+.11	-.08	-.02	-.08	+12.47	-1.49	-3.49	+1.17	+.03	-.01	+.00

TABLE 5.5. Event a ≤ k or a ≥ k+1, exact hypergeometric probability and relative tail error in per cent. for some normal, Poisson and binomial approximations. For distributions with N = 2r, probabilities and errors are the same for a ≤ k and a ≥ n-k, provided that Poisson approximations are applied to a ≤ n-k-1 when P[a ≤ k] must be evaluated for k ≥ n/2

Event	Probability	impr. (2.45)	square root (2.37)	(2.38)	χ (2.3)	u_h (2.1)	χ_t p.131	Poisson μ=nr/N	(3.11) λ_c	(3.9) λ_b	binomial $\tilde{r}$=r/N	(4.1) w	(4.10) w+γ	(4.9) w+δ
n = 5 r = 5 N = 20														
a ≤ 0	.1937	-2.50	+13.55	+1.42	-4.21	-1.05	-23.40	+47.92	+2.73	+2.90	+22.52	+1.45	+.10	+.02
a ≥ 2	.3661	+.24	-7.13	-.89	+4.56	+5.35	+20.39	-2.93	+4.15	+2.56	+.30	+.30	-.07	+.00
a ≥ 3	.0726	+.55	-6.02	+7.47	-6.34	+.67	+23.72	+81.11	+18.60	+18.87	+42.53	+5.21	+1.28	+.14
a ≥ 4	.0049	-2.35	+21.10	+54.57	-25.64	-9.08	+14.95	+680.69	+41.81	+77.76	+218.75	+23.13	+9.71	+1.44
n = 50 r = 50 N = 200														
a ≤ 6	.0092	+4.37	-.68	-9.15	+27.87	+29.77	+12.96	+273.75	+20.55	+1.11	+109.66	+5.31	-1.33	+.08
a ≤ 8	.0625	+.56	+3.96	-1.56	+5.09	+5.86	-4.21	+99.77	+8.74	-3.63	+46.48	+2.74	-.58	+.04
a ≤ 10	.2276	-.04	+3.99	+.65	-1.00	-.75	-7.10	+30.51	+2.16	-5.70	+15.19	+.97	-.17	+.01
a ≥ 13	.4936	+.02	-2.46	-.60	+1.29	+1.29	+5.10	-2.56	+.78	+5.71	-.94	+.01	+.01	+.00
a ≥ 15	.2234	+.02	-3.14	-.21	+.87	+1.13	+7.32	+23.08	+4.59	+13.89	+12.77	+1.16	-.07	+.02
a ≥ 17	.0678	+.30	-1.93	+2.26	-3.05	-2.34	+6.16	+92.81	+11.33	+28.12	+45.04	+3.60	-.07	+.07
a ≥ 19	.0134	+1.24	+2.67	+8.46	-11.64	-10.33	-.18	+287.42	+21.23	+51.52	+114.69	+7.82	+.18	+.16
n = 5 r = 8 N = 20														
a ≤ 0	.0511	+.60	+6.38	-.16	+11.43	+20.68	+.12	+164.93	-5.73	+12.43	+52.22	+3.60	+.05	+.07
a ≤ 1	.3065	-.54	+2.83	+.66	-2.42	-.90	-8.31	+32.46	-7.38	-.13	+9.94	+.74	+.05	+.01
a ≥ 3	.2962	-.05	-3.59	-.23	+.98	+2.55	+7.24	+9.16	+9.16	+9.16	+7.18	+.76	+.02	+.01
a ≥ 4	.0578	-1.04	-3.75	+4.79	-1.50	+6.67	+9.36	+147.23	+10.00	+42.40	+50.61	+4.62	+.65	+.11
a ≥ 5	.0036	-.93	+15.98	+32.46	+16.39	+41.39	+35.77	+1357.74	-12.13	+182.58	+183.50	+13.93	+3.36	+.54
n = 50 r = 80 N = 200														
a ≤ 12	.0054	+1.16	-.96	-3.36	+14.14	+16.17	+8.88	+617.06	-7.53	-2.25	+143.55	+7.17	-1.16	+.11
a ≤ 14	.0320	+.23	+.79	-1.01	+4.22	+5.29	+.42	+227.46	-7.14	-9.82	+68.48	+4.13	-.65	+.06
a ≤ 16	.1212	+.01	+1.34	+.03	+.41	+.90	-2.34	+82.44	-5.60	-13.48	+28.81	+1.96	-.29	+.03
a ≤ 18	.3102	-.01	+1.19	+.28	-.53	-.39	-2.41	+22.97	-3.54	-13.97	+8.20	+.58	-.08	+.01
a ≥ 21	.4317	+.00	-.98	-.22	+.48	+.52	+2.00	+2.12	+2.12	+15.74	+1.67	+.16	-.02	+.00
a ≥ 23	.2019	+.01	-1.23	-.01	+.21	+.50	+2.55	+38.37	+1.34	+33.07	+15.88	+1.26	-.13	+.02
a ≥ 25	.0674	+.02	-.85	+.95	-.95	-.23	+2.29	+132.43	-2.01	+62.90	+45.01	+3.25	-.27	+.05
a ≥ 27	.0156	+.04	+.71	+3.24	-3.03	-1.70	+1.11	+399.15	-9.04	+114.01	+101.27	+6.32	-.42	+.11
a ≥ 29	.0024	+.02	+4.18	+7.65	-5.71	-3.59	-.67	+1305.68	-20.35	+204.32	+211.87	+10.71	-.49	+.19

TABLE 5.6. Event $\underline{a} \leq k$ or $\underline{a} \geq k+1$, exact hypergeometric probability and relative tail error in per cent. for some normal, Poisson, and binomial approximations. For distributions with $N = 2r$, probabilities and errors are the same for $\underline{a} \leq k$ and $\underline{a} \geq n-k$, provided that Poisson approximations are applied to $\underline{a} \leq n-k-1$ when $P[\underline{a} \leq k]$ must be evaluated for $k \geq n/2$

Event	Probability	impr. (2.45)	square root (2.37)	square root (2.38)	χ (2.3)	u_h (2.1)	χ_t p.131	Poisson $\mu=nr/N$	(3.11) λ_c	(3.9) λ_b	binomial $\tilde{r}=r/N$	(4.1) w	(4.10) $w+\gamma$	(4.9) $w+\delta$
n = 5 r = 10 N = 20														
$\underline{a} \leq 0$	.0163	+1.74	+4.47	+5.99	+19.56	+35.61	+19.56	+405.02	-4.61	+31.43	+92.26	+6.69	+.66	+.17
$\underline{a} \leq 1$	.1517	-.67	-1.09	+.69	-.56	+3.53	-.56	+89.38	-10.42	+5.47	+23.60	+1.98	+.10	+.03
$\underline{a} \geq 3$	.5000	+.00	+.00	+.00	+.00	+.00	+.00	-8.76	+7.69	+4.64	+.00	+.00	+.00	+.00
n = 50 r = 100 N = 200														
$\underline{a} \leq 16$	.0026	+.21	+.91	+1.53	+5.96	+8.25	+5.96	+1354.11	-9.93	+.32	+195.59	+9.45	-.99	+.16
$\underline{a} \leq 18$	.0166	+.03	+.28	+.70	+1.84	+3.19	+1.84	+455.25	-9.81	-12.21	+95.79	+5.69	-.64	+.09
$\underline{a} \leq 20$	.0706	-.01	+.04	+.27	+.31	+1.02	+.31	+162.72	-8.20	-18.91	+43.51	+2.98	-.36	+.05
$\underline{a} \leq 22$	.2072	-.00	-.01	+.08	-.04	+.24	-.04	+53.25	-5.76	-21.17	+15.81	+1.18	-.15	+.02
$\underline{a} \leq 24$	.4352	-.00	-.00	+.01	-.01	+.02	-.01	+8.78	-3.21	-19.85	+1.99	+.15	-.02	+.00
n = 8 r = 8 N = 20														
$\underline{a} \leq 0$	.0039	+13.02	+4.04	+11.29	+51.21	+80.83	+43.40	+937.34	+22.86	+52.08	+327.44	+26.22	+2.52	+1.66
$\underline{a} \leq 1$	.0542	-.04	-3.44	+1.04	+4.39	+13.08	+.54	+215.71	-1.73	+12.67	+96.17	+10.78	+.86	+.57
$\underline{a} \leq 2$	.2596	-.27	-.72	-.02	-.95	+1.11	-3.25	+46.34	-8.28	-3.68	+21.49	+2.78	+.28	+.14
$\underline{a} \geq 4$	.3883	-.02	-1.74	-.29	+.42	+1.12	+2.26	+2.36	+9.97	+12.61	+4.53	+.92	+.05	+.04
$\underline{a} \geq 5$	.1132	-.29	-5.51	+.57	-.29	+4.99	+2.89	+93.73	+16.72	+45.70	+53.36	+8.23	+1.28	+.47
$\underline{a} \geq 6$	.0154	-.24	-4.68	+10.49	+4.02	+18.98	+8.96	+582.69	+7.28	+138.92	+222.58	+27.34	+6.39	+2.11
n = 80 r = 80 N = 200														
$\underline{a} \leq 23$	.0058	+.38	-.59	-.03	+5.35	+7.23	+3.60	+947.09	+5.74	-29.60	+321.21	+23.77	-.64	+1.05
$\underline{a} \leq 25$	.0272	+.10	-.31	-.10	+1.92	+3.05	+.55	+351.26	+.12	-34.35	+147.86	+14.32	-.35	+.65
$\underline{a} \leq 27$	.0921	+.01	+.02	-.02	+.37	+.97	-.69	+134.75	-2.57	-35.63	+65.11	+7.58	-.15	+.35
$\underline{a} \leq 29$	.2310	-.01	+.25	+.06	-.15	+.10	-.92	+46.30	-3.26	-33.97	+23.82	+3.08	-.04	+.14
$\underline{a} \leq 31$	.4422	-.00	+.31	+.08	-.18	-.15	-.71	+7.74	-2.73	-29.73	+3.48	+.42	-.01	+.02
$\underline{a} \geq 34$	.3288	-.00	-.51	-.09	+.15	+.27	+.80	+17.09	+3.44	+47.80	+10.61	+1.59	+.01	+.08
$\underline{a} \geq 36$	.1512	+.00	-.75	+.04	-.01	+.39	+.90	+73.28	+4.06	+89.76	+39.83	+5.39	+.09	+.26
$\underline{a} \geq 38$	.0528	+.02	-.73	+.57	-.36	+.47	+.85	+212.34	+2.99	+162.52	+99.66	+11.67	+.32	+.57
$\underline{a} \geq 40$	.0137	+.08	-.13	+1.85	-.76	+.66	+.75	+599.55	-.92	+292.46	+225.59	+21.23	+.82	+1.03
$\underline{a} \geq 42$	.0026	+.19	+1.53	+4.38	-.96	+1.21	+.86	+1878.12	-8.62	+536.09	+511.42	+35.19	+1.75	+1.69

Table 5.7. Event $\underline{a} \leq k$ or $\underline{a} \geq k+1$, exact hypergeometric probability and relative tail error in per cent. for some normal, Poisson and binomial approximations. For distributions with $N = 2r$, probabilities and errors are the same for $\underline{a} \leq k$ and $\underline{a} \geq n-k$, provided that Poisson approximations are applied to $\underline{a} \leq n-k-1$ when $P[\underline{a} \leq k]$ must be evaluated for $k \geq n/2$

Event	Probability	impr. (2.45)	square root (2.37)	square root (2.38)	χ (2.3)	u_h (2.1)	χ_t p.131	Poisson $\mu=nr/N$	(3.11) λ_c	(3.9) λ_b	binomial $\tilde{r}=r/N$	(4.1) w	(4.10) $w+\gamma$	(4.9) $w+\delta$
n = 8 r = 10 N = 20														
$\underline{a} \leq 1$	.0099	+.50	-3.51	+11.03	+13.72	+32.15	+13.72	+826.59	+4.61	+48.40	+255.71	+25.89	+4.30	+1.78
$\underline{a} \leq 2$	.0849	-.42	-5.20	+.69	+.65	+7.18	+.65	+180.45	-11.68	+5.69	+70.23	+9.29	+1.13	+.51
$\underline{a} \leq 3$	.3250	-.09	-1.41	-.19	-.28	+1.00	-.28	+33.39	-12.53	-9.66	+11.79	+1.76	+.17	+.09
n = 80 r = 100 N = 200														
$\underline{a} \leq 29$	.0012	+.07	+.99	+3.34	+4.15	+6.80	+4.15	+3595.72	-3.72	-42.17	+682.94	+38.41	+.52	+1.74
$\underline{a} \leq 31$	.0070	+.01	-.18	+1.40	+1.69	+3.44	+1.69	+1130.30	-7.97	-49.25	+307.59	+24.52	+.20	+1.13
$\underline{a} \leq 33$	.0301	-.00	-.53	+.45	+.52	+1.59	+.52	+402.26	-9.24	-52.27	+141.56	+14.55	+.06	+.68
$\underline{a} \leq 35$	.0969	-.00	-.45	+.08	+.07	+.65	+.07	+150.18	-8.46	-52.16	+62.19	+7.60	+.00	+.36
$\underline{a} \leq 37$	.2353	-.00	-.23	-.01	-.02	+.21	-.02	+50.72	-6.43	-49.24	+22.49	+3.05	-.01	+.14
$\underline{a} \leq 39$	.4426	-.00	-.03	-.00	-.01	+.03	-.01	+8.21	-3.95	-43.62	+2.91	+.41	-.00	+.02
n = 10 r = 10 N = 20														
$\underline{a} \leq 2$	.0115	-.04	-5.36	+10.14	+10.14	+27.31	+10.14	+983.26	+19.53	+49.24	+375.25	+47.80	+10.61	+5.15
$\underline{a} \leq 3$	.0894	-.35	-5.72	+.46	+.46	+6.76	+.46	+196.29	-8.59	-1.41	+92.15	+16.13	+3.01	+1.50
$\underline{a} \leq 4$	.3281	-.07	-1.51	-.24	-.24	+1.01	-.24	+34.24	-13.13	-16.95	+14.88	+2.99	+.50	+.26
n = 100 r = 100 N = 200														
$\underline{a} \leq 39$	.0014	+.03	+.76	+3.24	+3.24	+5.76	+3.24	+4374.94	+20.76	-70.02	+1119.75	+74.58	+6.58	+5.35
$\underline{a} \leq 41$	.0080	+.00	-.34	+1.32	+1.32	+3.00	+1.32	+1303.81	+6.85	-73.04	+453.99	+45.67	+4.14	+3.48
$\underline{a} \leq 43$	.0329	-.00	-.62	+.40	+.40	+1.44	+.40	+447.11	-.74	-73.58	+194.17	+26.29	+2.42	+2.10
$\underline{a} \leq 45$	.1015	-.00	-.50	+.05	+.05	+.61	+.05	+162.95	-4.21	-71.98	+81.40	+13.47	+1.25	+1.11
$\underline{a} \leq 47$	.2398	-.00	-.25	-.02	-.02	+.21	-.02	+54.16	-4.96	-68.19	+28.71	+5.36	+.50	+.45
$\underline{a} \leq 49$	.4438	-.00	-.04	-.01	-.01	+.03	-.01	+8.43	-4.10	-61.87	+3.70	+.72	+.07	+.06

REFERENCES

In this list one finds the section number(s) in which a reference is mentioned. The fourth reference ANSCOMBE (1948) e.g. is found in section 3c of Chapter II and III. Many people have helped the author by calling his attention to certain publications. Special thanks are due to Dr. A.J. MAYNE who mentioned ANSCOMBE's work on the 2/3 power transformation.

ANDERSON, T.W., & H. BURSTEIN (1967), Approximating the upper binomial confidence limit, J. Amer. Stat. Assoc. 62, 857-862; Summary.

ANDERSON, T.W., & H. BURSTEIN (1968), Approximating the lower binomial confidence limit, J. Amer. Stat. Assoc. 63, 1413-1416; Summary.

ANDERSON, T.W., & S.M. SAMUELS (1967), Some inequalities among binomial and Poisson probabilities, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, Univ. of California Press; III.6.

ANSCOMBE, F.J. (1948), The transformation of Poisson, binomial and negative binomial data, Biometrika 35, 246-254; II.3c, III.3c.

ANSCOMBE, F.J. (1953), Discussion on the paper by H. Hotelling, New light on the correlation coefficient and its transforms, J. Roy. Stat. Soc. Series B 15, 229-230; II.4a.

ANSCOMBE, F.J. (1960), Notes on sequential sampling plans, J. Roy. Stat. Soc. Series A 123, 297-306; II.4a.

BAHADUR, R.R. (1960), Some approximations to the binomial distribution function, Ann. Math. Stat. 31, 43-54; I.1.

BARTLETT, M.S. (1936), The square root transformation in analysis of variance, J. Roy. Stat. Soc. Suppl. 3, 68-78; II.3c.

BLOM, G. (1954), Transformations of the binomial, negative binomial, Poisson and χ^2 distributions, Biometrika 41, 302-316; correction note, Biometrika 43 (1956), 235; I.4, II.3c, II.4a, II.4b, III.4b.

BOLSHEV, L.N., B.V. GLADKOV & M.V. SHCHEGLOVA (1961), Tables for the calculation of B- and z-distribution functions, Theory of Prob. and its Applic. 6, 410-419 (English translation of Teoriya Veroyatnostei i ee Primenenyia); Summary, I.3, III.5b, III.6.

BOLSHEV, L.N. (1963), Asymptotically Pearson transformations, Theory of Prob. and its Applic. 8, 121-146 (English translation of Teoryia Veroyatnostei i ee Primeneniya); III.6.

BOLSHEV, L.N. (1964), Some applications of Pearson transformations, Review of the Internat. Stat. Institute 32, 14-16; III.6.

BOLSHEV, L.N. (1964a), Distributions related to the hypergeometric distribution, Theory of Prob. and its Applic. 9, 619-624 (English translation of Teoryia Veroyatnostei i ee Primeneniya); IV.4.

BORGES, R. (1970), Eine Approximation der Binomialverteilung durch die Normalverteilung der Ordnung 1/n, Zeit. Wahrsch. verw. Geb. 14, 189-199; I.3, I.4, III.3c, III.4b.

BOSCHLOO, R.D. (1970), Raised conditional level of significance for the 2 x 2 table when testing the equality of two probabilities, Stat. Neerl. 24, 1-35; IV.1.

CAMP, B.H. (1951), Approximation to the point binomial, Ann. Math. Stat. 22, 130-131; III.4a.

CAMPBELL, G.A. (1923), Probability curves showing Poisson's exponential summation, Bell System Techn. J. 2, 95-113; also in: Collected Papers, American Telephone and Telegraph Cy., New York 1937, 224-242; II.1b, III.6.

CHENG, T.T. (1949), The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan, Bull. Amer. Math. Soc. 55, 396-401; II.4b.

CORNISH, E.A.,& R.A. FISHER (1937), Moments and cumulants in the specification of distributions, Review of the Internat. Stat. Inst. 5, 307-320; Reprinted in: R.A. Fisher, Contrib. to Math. Stat., Wiley, New York 1950; I.4, II.2, III.2.

CURTISS, J.H. (1943), On transformations used in the analysis of variance, Ann. Math. Stat. 14, 107-122; II.3c, III.3c, III.4b.

EEDEN, C. van, & J.TH. RUNNENBURG (1960), Conditional limit-distributions for the entries in a 2 x 2 table, Stat. Neerl. 14, 111-126; IV.1, IV.3.

ESSEEN, C.G. (1945), Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law, Acta Math. 77, 1-125; II.2, III.2.

FELLER, W. (1945), On the normal approximation to the binomial distribution, Ann. Math. Stat. 16, 319-329; corrections in Ann. Math. Stat. 21 (1950), 302; III.3a, III.4c, IV.2e.

FELLER, W. (1957), An introduction to probability theory and its applications, 2nd edition, Wiley, New York; the quoted approximation is not found in the 3rd edition (1968); III.3a.

FIELLER, E.C. (1932), The distribution of the index in a normal bivariate population, Biometrika 24, 428-440; III.4a.

FREEMAN, M.F.,& J.W. TUKEY (1950), Transformations related to the angular and the square root, Ann. Math. Stat. 21, 607-611; Summary, II.3c, III.3c, III.4c.

GEARY, R.C. (1930), The frequency distribution of the quotient of two normal variates, J. Roy. Stat. Soc. Series A 93, 442-446; III.4a.

GEBHARDT, F. (1969), Some numerical comparisons of several approximations to the binomial distribution, J. Amer. Stat. Assoc. 64, 1638-1646; I.4.

GENERAL ELECTRIC COMPANY (1962), Tables of the Individual and Cumulative Terms of Poisson Distribution, Van Nostrand, Princeton (N.J.); II.1b.

HALD, A. (1952), Statistical Theory with Engineering Applications, Wiley, New York; III.3c.

HALD, A. (1952a), Statistical tables and formulas, Wiley, New York; III.3c.

HALD, A. (1967), Determination of single sampling attribute plans with given producer's and consumer's risk, Technometrics 9, 401-415; IV.4.

HARTER, H.L. (1964), New tables of the incomplete gamma function ratio and of percentage points of the chi-square and beta distributions, Aerospace Research laboratories; for sale by the Superintendent of Documents, U.S. Government Printing Office, Washington D.C. 20402; II.1b.

HARVARD (1955), Tables of the cumulative binomial probability distribution, Harvard Univ. Press, Cambridge (Mass.); III.1.

HEMELRIJK, J. (1962), De asymptotische eigenschappen van de binomiale verdeling, Stat. Neerl. 16, 255-258; III.1.

HEMELRIJK, J. (1967), The hypergeometric, the normal and chi-squared, Stat. Neerl. 21, 225-231; IV.2d.

HILL, G.W., & A.W. DAVIS (1968), Generalized asymptotic expansions of Cornish-Fisher type, Ann. Math. Stat. 39, 1264-1273; II.2, IV.2.

HODGES Jr., J.L., & L. LECAM (1960), The Poisson approximation to the Poisson binomial distribution, Ann. Math. Stat. 31, 737-740; I.4.

JONGE, H. de (1960), Inleiding tot de medische statistiek, deel 2, Ned. Inst. voor Praeventieve Geneeskunde, Leiden; III.3c.

KHAMIS, S.H., in cooperation with W. RUDERT (1965), Tables of the Incomplete Gamma Function Ratio, Justus von Liebig, Darmstadt; II.1b.

LAUBSCHER, N.F. (1960), On the stabilization of the Poisson variance, Trab. de Estadistica 11, 199-207; II.3c.

LAUBSCHER, N.F. (1961), On stabilizing the binomial and negative binomial variances, J. Amer. Stat. Assoc. 56, 143-150; III.3c.

LECAM, L. (1960), An approximation theorem for the Poisson binomial distribution, Pacific J. Math. 10, 1181-1197; I.4.

LEHMANN, E.L. (1959), Testing statistical hypotheses, Wiley, New York; IV.1.

LIEBERMAN, G.J., & D.B. OWEN (1961), Tables of the hypergeometric probability distribution, Stanford Studies in Math. and Stat. III, Stanford Univ. Press, Stanford (Cal.); IV.1.

MAKABE, H., & H. MORIMURA (1955), A normal approximation to Poisson distribution, Rep. Stat. Appl. Res. Japanese Union of Scientists and Engineers 4, 37-46; II.4b.

MOLENAAR, W. (1969), How to poison Poisson (when approximating binomial tails), Stat. Neerl. 23, 19-40 and 241; III.6.

MOLENAAR, W. (1969a), Normal approximations to the Poisson distribution, to appear in: Random Counts in Scientific Work, vol. 2, edited by G.P. Patil, Pennsylvania State University Press. Preliminary version: Report S 412, Statistische Afdeling, Mathematisch Centrum, Amsterdam 1969; Summary.

MOLINA, E.C. (1945), Poisson's Exponential Binomial Limit, Van Nostrand, New York; II.1b.

MOORE, P.G. (1957), Transformations to normality using fractional powers of the variable, J. Amer. Stat. Assoc. 52, 237-246; II.4a.

MOSTELLER, F., & J.W. TUKEY (1949), The uses and usefulness of binomial probability paper, J. Amer. Stat. Assoc. 44, 174-212; III.3c.

NICHOLSON, W.L. (1956), On the normal approximation to the hypergeometric distribution, Ann. Math. Stat. 27, 471-483; IV.2d, IV.2e.

ORD, J.K. (1968), Approximations to distribution functions which are hypergeometric series, Biometrika 55, 243-248; IV.2d, IV.4.

ORDNANCE CORPS (1952), Tables of the Cumulative Binomial Probabilities, Document ORDP 20-1 of the Ordnance Corps of the U.S. Army; III.1.

PAULSON, E. (1942), An approximate normalization of the analysis of variance distribution, Ann. Math. Stat. 13, 233-235; III.4a.

PEARSON, K. (1922), Tables of the Incomplete Gamma Function, H.M. Stationery Office, London; II.1b.

PEARSON, K. (1934), Tables of the Incomplete Beta Function, Cambridge Univ. Press for the Biometrika Trustees, Cambridge; III.1.

PEARSON, E.S.,& H.O. HARTLEY (1954), Biometrika Tables for Statisticians, vol.1, Cambridge University Press, Cambridge; II.1b.

PEIZER, D.B.,& J.W. PRATT (1968), A normal approximation for binomial, F, beta and other common, related tail probabilities, J. Amer. Stat. Assoc. 63, 1417-1456 (part I) and 1457-1483 (part II, by J.W. Pratt alone); I.2, I.3, I.4, II.2, II.5, III.2, III.5b.

PROHOROV, YU.V. (1953), Asymptotic behavior of the binomial distribution, Uspehi Mat. Nauk (N.S.) 8 (1953) no. 3(35), 135-142 (Russian); English Translation in Selected Translations in Math. Stat. and Probability, vol. 1, Amer. Math. Soc., Providence (R.I.) 1961, 87-96; I.4.

RAFF, M.S. (1955), The comparative accuracy of several approximations to the cumulative binomial distribution, M.A. Thesis, The American University, Washington D.C.; I.4, III.3c, III.4c, III.6.

RAFF, M.S. (1956), On approximating the point binomial, J. Amer. Stat. Assoc. 51, 293-303; I.4, III.3c, III.4a, III.4c, III.6.

RIET, R.P. van de (1968), Formula Manipulation in Algol 60, M.C. Tract no. 17, Mathematisch Centrum, Amsterdam; I.2.

RIORDAN, J. (1949), Inversion formulas in normal variable mapping, Ann. Math. Stat. 20, 417-425; II.2.

SANDIFORD, P.J. (1960), A new binomial approximation for use in sampling from finite populations, J. Amer. Stat. Assoc. 55, 718-722; IV.4.

SHANNON, S. (1938), An alternative method of solution of certain fundamental problems in the theory of risk, Rec. Amer. Inst. Actuar. 27, 372-397; III.3a.

SHANNON, S.,& C.A. SPOERL (1942), Comparative aspects of the point binomial polygon and its associated normal curve of error, Rec. Amer. Inst. Actuar. 31, 208-226; III.3a.

VERENIGING VOOR STATISTIEK, Statistische Tabellen en Nomogrammen, Stenfert Kroese, Leiden; II.1b.

WEINTRAUB, S. (1963), Tables of the Cumulative Binomial Probability Distribution for Small Values of p, the Free Press of Glencoe, New York; III.1.

WILSON, E.B.,& M.M. HILFERTY (1931), The distribution of chi-square, Proc. Nat. Acad. Sci. 17, 684-688; II.3b, II.4a, III.4a.

WISE, M.E. (1946), The use of the negative binomial distribution in an industrial sampling problem, J. Royal Stat. Soc. Supplement 8, 202-211; III.6.

WISE, M.E. (1950), The incomplete Beta function as a contour integral and a quickly converging series for its inverse, Biometrika 37, 208-218; III.6.

WISE, M.E. (1954), A quickly convergent expansion for cumulative hypergeometric probabilities, direct and inverse, Biometrika 41, 317-329; Summary, IV.2c, IV.4.

WISHART, J. (1956), χ^2 probabilities for large numbers of degrees of freedom, Biometrika 43, 92-95; II.4a.

WISHART, J. (1957), An approximate formula for the cumulative z-distribution, Ann. Math. Stat. 28, 504-510; III.5b.

ZWET, W.R. van (1964), Convex transformations of random variables, M.C. Tract no. 7, Mathematisch Centrum, Amsterdam; III.2.

LIST OF SYMBOLS

Underlined symbols denote random variables. In particular $\underline{a}$ has a hypergeometric, $\underline{x}$ has a Poisson and $\underline{y}$ has a binomial distribution. We put

$E\,\underline{z}$ = expectation of a random variable $\underline{z}$;

$F_\lambda(k) = P[\underline{x} \leq k]$ = Poisson distribution function, I (1.1), II (1.1);

$G_{n,p}(k) = P[\underline{y} \leq k]$ = binomial distribution function, I (1.2), III (1.1);

$H_{n,r,N}(k) = P[\underline{a} \leq k]$ = hypergeometric distr. function, I (1.3), IV (1.1);

k is exclusively used as the argument of F_λ, $G_{n,p}$ or $H_{n,r,N}$;

$m = N-n$, n, r, $s = N-r$ are the hypergeometric marginals;

n and p are the binomial parameters;

N is the grand total of the 2×2 table (hypergeometric parameter);

O and o are the LANDAU-BACHMANN order symbols, cf. section I.2;

$P[\ldots]$ denotes the probability of an event;

λ is the Poisson parameter;

λ_0 is the exact parameter of a Poisson approximation, cf.III.6 and IV.3;

$\mu = nrN^{-1} = E\,\underline{a}$ (hypergeometric distribution);

ξ is the exact deviate of a normal approximation, II.2, III.2, IV.2;

$\sigma^2(\underline{z})$ is the variance of a random variable $\underline{z}$;

σ^2 without argument denotes the binomial variance: $\sigma^2 = npq$;

$\tau^2 = mnrsN^{-3}$ (hypergeometric distribution);

ϕ is the standard normal density function, I (2.2);

Φ is the standard normal distribution function , I (2.1).