SA

DUPLICAAT

stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 9/73 NOVEMBER

W. MOLENAAR

SIMPLE APPROXIMATIONS TO THE POISSON,
BINOMIAL AND HYPERGEOMETRIC DISTRIBUTIONS

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

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

1st printing: august 1971
2nd printing: november 1973

SIMPLE APPROXIMATIONS TO THE POISSON BINOMIAL AND HYPERGEOMETRIC DISTRIBUTIONS

by W. Molenaar

Mathematisch Centrum, Amsterdam, and
Pennsylvania State University *)

ABSTRACT

The classical approximations to some discrete distributions are almost universally used. However, more accurate results can be obtained without any substantial increase of computational effort. This paper presents a set of recommendations for improved approximations to cumulative probabilities, confidence bounds and quantiles of three well known distributions.

Quick evaluation, using a minimum of tables and at most a slide rule or desk calculator, is often far more desirable than consultation of a specialized table in a library, or than waiting for access to an electronic computer. This asks for the choice of an approximation achieving a maximum of accuracy with a minimum of work.

The present paper is an effort to solve this problem. It intends to give enough information for people just desiring to use the results. References will be mentioned for readers interested in a more detailed discussion to the topic.

----

*) Now at the Rijksuniversiteit in Groningen (Netherlands)

1. INTRODUCTION

The following example is typical for this paper. The probability of 8 or more successes in 20 independent trials with success probability .2 at each trial, equals .032 according to binomial tables. Without immediate access to such tables, or to an electronic computer, most statisticians would evaluate the standard normal distribution function at the argument $(8-\frac{1}{2}-4)/(3.2)^{\frac{1}{2}}$, and find .019, or perhaps they would obtain .051 from the Poisson distribution with parameter np=4. However, without substantial increase of effort they might have found .033 from the "square root normal" or .035 from the "Bolshev Poisson" approximation, see (2.2) and (2.5) below, and both values are much closer to the actual value of .032 than the classical results. After a more time consuming computation, (2.7) or (2.8) produces a result within .0002 of the true value.

Although the errors depend in a complicated way on the values of parameters and probability, the general trend of the quoted figures is found in a large majority of the applications. Our choice of simple but accurate approximations, partly previously published and partly new, is based on asymptotic expansions and numerical investigations [5,6,7].

The recommendations in the following Tables are kept as simple as possible. Obviously we were forced to neglect some exceptional cases, and to use some personal judgement in evaluating accuracy and computational labor. Although the recommended formulae are usually rather superior, there will always remain some surprising parameter values for which a generally bad approximation is suddenly very accurate.

This study intends to give improved methods for hand or desk calculation, with a minimum of tables. On an electronic computer, exact evaluation is possible unless the parameter values are extreme, and when a computer programmer uses approximations he would not share our standards of accuracy and labor.

Tables 1-8 constitute the core of this paper. We comment on these Tables by means of footnotes, and close the section by an example using Tables 5 and 6.

TABLE 1 : CUMULATIVE POISSON PROBABILITIES

Wanted: probability of k *or less* [1] Poisson distributed events when λ are expected, i.e.

$$(1.1) \qquad P[X\leq k|\lambda] = \sum_{j=0}^{k} e^{-\lambda} \lambda^j/j!$$

Exact evaluation of (1.1) is satisfactory for small λ (say $\lambda < .1$ or may be $\lambda < .5$) and more generally when $k = 0$.

Quick work normal approximation [2]:

$$(1.2) \qquad \Phi(2\{k+1\}^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}) \quad \text{for tails } ^{3)};$$

$$(1.3) \qquad \Phi(2\{k+\tfrac{3}{4}\}^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}})$$

for probabilities between .06 and .94 (or .05 and .93 for roughly $\lambda < 15$).

Never use $\Phi((k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}})$ or $\Phi((k-\lambda)\lambda^{-\frac{1}{2}})$, as (1.2) and (1.3) are essentially better and just as simple (cf. Appendix B).

Accurate normal approximations [2]:

$$(1.4) \qquad \Phi(2\{k + (t+4)/9\}^{\frac{1}{2}} - 2\{\lambda + (t-8)/36\}^{\frac{1}{2}}),$$

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

$$(1.5) \qquad \Phi(\{k - \lambda + \frac{2}{3} + \frac{.022}{k+1}\} \{1 + g(\frac{k+\frac{1}{2}}{\lambda})\}^{\frac{1}{2}} \lambda^{-\frac{1}{2}}),$$

with $g(z) = (1 - z^2 + 2z \log z)(1-z)^{-2}$ tabled in [9].

EXAMPLE: $\lambda = 10$, $k = 4$. Exact value (1.1) $P[X\leq 4|\lambda=10] = .029$. Substitution of $\lambda = 10$ and $k = 4$ in (1.2) gives $\Phi(2\sqrt{5}-2\sqrt{10}) = \Phi(-1.852) = .032$, with a relative error of $(.032-.029)/.029$ which is 9%. Similarly, the relative error is found to be 40% for the classical $\Phi((k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}})$, -16% for (1.3), -.02% for (1.4) and -.04% for (1.5). The latter two can safely be used even for λ as small as .5, whereas (1.2) and (1.3) are rather rough unless λ exceeds 10.

For footnotes consult section 1.

TABLE 2 : CUMULATIVE BINOMIAL PROBABILITIES

Wanted: *probability of* k *or less* [1] successes in n independent trials with success probability p, i.e.

(2.1) $$P[X\leq k|n,p] = \sum_{j=0}^{k} \binom{n}{j} p^j q^{n-j}, \quad \text{where } q = 1-p.$$

Quick work approximation [2]:

(2.2) $$\Phi(2\{k+1\}^{\frac{1}{2}}q^{\frac{1}{2}} - 2\{n-k\}^{\frac{1}{2}}p^{\frac{1}{2}}) \quad \text{for tails [3];}$$

(2.3) $$\Phi(\{4k+3\}^{\frac{1}{2}}q^{\frac{1}{2}} - \{4n-4k-1\}^{\frac{1}{2}}p^{\frac{1}{2}}) \quad \text{for probabilities between 0.5 and .93.}$$

When p is close to .5, say $.25 \leq p \leq .75$ when n = 3 or $.40 \leq p \leq .60$ when n = 30 or $.46 \leq p \leq .54$ when n = 300, it is better to use (2.3) for tails [3], and for probabilities between .06 and .94 to use

(2.4) $$\Phi(\{4k+2\tfrac{1}{2}\}^{\frac{1}{2}}q^{\frac{1}{2}} - \{4n-4k-1\tfrac{1}{2}\}^{\frac{1}{2}}p^{\frac{1}{2}}).$$

When *cumulative Poisson tables* are available and p is small [5], say $p \leq .4$, when n = 3 or $p \leq .3$ when n = 30 or $p \leq .2$ when n = 300, use [4]

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

Accurate approximations [2]: use for $p = \frac{1}{2}$

(2.6) $$\Phi(\{2k+2+b\}^{\frac{1}{2}} - \{2n-2k+b\}^{\frac{1}{2}}) \quad \text{where } b = \frac{(2k+1-n)^2-10n}{12n},$$

and for any other p, consulting (1.5) for the function g,

(2.7) $$\Phi(\{k + \frac{2}{3} - (n+\frac{1}{3})p\} \{1 + qg(\frac{k+\frac{1}{2}}{np}) + pg(\frac{n-k-\frac{1}{2}}{nq})\}^{\frac{1}{2}} \{(n+\frac{1}{6})pq\}^{-\frac{1}{2}}).$$

When *cumulative Poisson tables* are available and p is small [5], say $p \leq .4$ when n = 3 or $p \leq .24$ when n = 30 or $p \leq .12$ when n = 300, use [4]

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

For footnotes consult section 1.

TABLE 3 : CUMULATIVE HYPERGEOMETRIC PROBABILITIES

Wanted: *probability of* k *or less* [1] red balls in n drawings without replacement from N balls of which r are red, i.e.

$$(3.1) \qquad \sum_{j=0}^{k} \binom{r}{j}\binom{N-r}{n-j}/\binom{N}{n} \qquad \text{(assumption: } n \le r \le N/2)\ ^{6)}.$$

When *cumulative binomial tables* can be applied, for quick work

$$(3.2) \qquad \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \qquad \text{with } p = (2r-k)/(2N-n+1);$$

for accurate results even when n/N > .1,

$$(3.3) \qquad \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \qquad \text{with } p = \frac{2r-k}{2N-n+1} - \frac{n(2k+1-2nrN^{-1})}{3(2N-n+1)^2}.$$

Without binomial tables, use [2]

$$(3.4) \qquad \Phi(2\{N-1\}^{-\frac{1}{2}}\{(k+1)^{\frac{1}{2}}(N-n-r+k+1)^{\frac{1}{2}} - (n-k)^{\frac{1}{2}}(r-k)^{\frac{1}{2}}\}) \quad \text{for tails } ^{3)};$$

$$(3.5) \qquad \Phi(2N^{-\frac{1}{2}}\{(k+\tfrac{3}{4})^{\frac{1}{2}}(N-n-r+k+\tfrac{3}{4})^{\frac{1}{2}} - (n-k-\tfrac{1}{4})^{\frac{1}{2}}(r-k-\tfrac{1}{4})^{\frac{1}{2}}\})$$

for probabilities between .05 and .93.

When *cumulative Poisson tables* are available but binomial tables are not, one may use

$$(3.6) \qquad \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! \quad \text{with } \lambda = \tfrac{1}{2}(2n-k)(2r-k)/(2N-n-r+1) \quad \text{for } n/N \le r/N \le .1;$$

$$(3.7) \qquad \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j! \quad \text{with } \lambda = \mu + (\mu-k)(2r-n+10\mu)/(3N), \quad \text{where } \mu = nr/N, \text{ otherwise};$$

a somewhat more accurate but also more cumbersome choice is

$$(3.9) \qquad \lambda = \frac{(12n-2np-7k)np}{12n-8np-k+k/n} \quad \text{where } p = \frac{2r-k}{2N-n+1}.$$

For footnotes consult section 1.

TABLE 4 : POISSON CONFIDENCE BOUNDS

Wanted: *upper bound* $\lceil\lambda$ for the Poisson parameter λ with (one-sided) confidence coefficient [7] $1-\alpha$, when c events have been observed, i.e. solution of

$$(4.1) \qquad \sum_{j=0}^{c} e^{-\lceil\lambda} \lceil\lambda^j/j! = \alpha,$$

or *lower bound* $\lfloor\lambda$ similarly found from

$$(4.2) \qquad \sum_{j=c}^{\infty} e^{-\lfloor\lambda} \lfloor\lambda^j/j! = \alpha.$$

Exact solution: when $c = 0$, $\lceil\lambda = -\log \alpha$ and $\lfloor\lambda = 0$, when $c = 1$. $\lfloor\lambda = -\log(1-\alpha)$.

Simple normal approximation [2], for ξ consult our Table 8:

$$(4.3) \qquad \lceil\lambda \approx (\{c+1\}^{\frac{1}{2}} + \tfrac{1}{2}\xi)^2;$$

$$(4.4) \qquad \lfloor\lambda \approx (c^{\frac{1}{2}} - \tfrac{1}{2}\xi)^2.$$

Better normal approximation [2], for ξ consult our Table 8:

$$(4.5) \qquad \lceil\lambda \approx \{c+1\}\ \{1 + \xi/(3\{c+1\}^{\frac{1}{2}}) - 1/(9\{c+1\})\}^3;$$

$$(4.6) \qquad \lfloor\lambda \approx c\{1 - \xi/(3c^{\frac{1}{2}}) - 1/(9c)\}^3.$$

Error bounds, uniformly for $.005 \leq \alpha \leq .2$. The *relative* error in the confidence bound value is less than 1%: for (4.3) when $c > 11$, for (4.4) when $c \geq 34$, for (4.5) for all $c \geq 0$ and for (4.6) when $c \geq 8$. The *absolute* difference between the exact bound and (4.5) or (4.6) rarely exceeds .02 and never exceeds .041.

EXAMPLE: Lower bound for λ with .99 confidence after observing $c = 10$. From (4.4) one finds, inserting $c = 10$ and $\xi = 2.326$,
$\lfloor\lambda \approx (3.162-1.163)^2 = 3.996$, but (4.6) leads to
$\lfloor\lambda \approx 10\{1 - .775/3.162 - 1/90\}^3 = 4.132$, whereas the true value, solution of (4.2), equals 4.129.

For footnotes consult section 1.

TABLE 5 : BINOMIAL CONFIDENCE BOUNDS

Wanted: *upper bound* $\lceil p$ for the binomial parameter p with (one-sided) confidence coefficient [7] $1-\alpha$, when c successes in n trials have been observed, or similar *lower bound* $\lfloor p$, i.e. solutions of

(5.1) $$\sum_{j=0}^{c} \binom{n}{j} \cdot \lceil p^{j}(1-\lceil p)^{n-j} = \alpha \quad \text{and} \quad \sum_{j=c}^{n} \binom{n}{j} \lfloor p^{j}(1-\lfloor p)^{n-j} = \alpha .$$

Let $\lceil\lambda$ and $\lfloor\lambda$ be the Poisson confidence bounds for the same α and c, looked up in tables or obtained from our Table 4.

For quick work, with Poisson bounds available:

(5.2) $$\lceil p \approx 2\lceil\lambda/(2n+\lceil\lambda-c) \qquad \text{when } c/n < .3;$$

(5.3) $$\lfloor p \approx 2\lfloor\lambda/(2n+\lfloor\lambda-c-1) \qquad \text{when } c/n < .5;$$

use $\lceil p(c) = 1 - \lfloor p(n-c)$ when $c/n > .5$ and $\lfloor p(c) = 1 - \lceil p(n-c)$ when $c/n > .7$, applying (5.2) and (5.3) with c replaced by n-c.

Normal approximation [2], satisfactory for roughly $\lceil p$ and $\lfloor p$ between .3 and .7 when $n \geq 10$, between .05 and .95 when $n \geq 60$:

(5.4) $$\lceil p \approx [c + \psi + \xi\{(c+1-\omega)(n-c-\omega)/(n+11\omega-4)\}^{\frac{1}{2}}]/(n+2\psi-1);$$

(5.5) $$\lfloor p \approx [c - 1 + \psi - \xi\{(c-\omega)(n+1-c-\omega)/(n+11\omega-4)\}^{\frac{1}{2}}]/(n+2\psi-1);$$

consult Table 8 for ξ, ψ and ω.

Accurate approximation [8] based on Poisson bounds:

(5.6) $$\lceil p \approx \lceil\lambda\{1 - (\lceil\lambda-2)/(4n)\}/\{n + (\lceil\lambda-2c-2)/4\} \qquad \text{for } c < \tfrac{1}{2}n-2;$$

(5.7) $$\lfloor p \approx \lfloor\lambda\{1 - (\lfloor\lambda+2)/(3n)\}/\{n + (\lfloor\lambda-3c-1)/6\} \qquad \text{for } c < \tfrac{1}{2}n+2;$$

for larger c use $\lceil p(c) = 1 - \lfloor p(n-c)$ and $\lfloor p(c) = 1 - \lceil p(n-c)$, evaluating (5.6) or (5.7) with c replaced by n-c with its Poisson bounds. As long as $\lceil p$ or $\lfloor p$ lie between .05 and .95, (5.6) and (5.7) remain satisfactory when the Poisson bounds $\lceil\lambda$ and $\lfloor\lambda$ are replaced by their approximations given in Table 4.

For footnotes and an example consult section 1.

TABLE 6 : HYPERGEOMETRIC CONFIDENCE BOUNDS

Wanted: *upper bound* $\lceil r$ for the hypergeometric parameter r with (one-sided) confidence coefficient [7] $1-\alpha$, when c red balls have been observed in a sample of size [6] $n \leq \frac{1}{2}N$, drawn without replacement from N balls out of which r are red, i.e. solution of

(6.1) $$\sum_{j=0}^{c} \binom{\lceil r}{j}\binom{N-\lceil r}{n-j}/\binom{N}{n} = \alpha,$$

or similar *lower bound* $\lfloor r$, i.e. solution of

(6.2) $$\sum_{j=c}^{n} \binom{\lfloor r}{j}\binom{N-\lfloor r}{n-j}/\binom{N}{n} = \alpha.$$

Accurate approximation based on binomial bounds:

(6.3) $$\lceil r \approx \tfrac{1}{2}c + \tfrac{1}{2}(2N-n+1)\lceil p;$$

(6.4) $$\lfloor r \approx \tfrac{1}{2}(c-1) + \tfrac{1}{2}(2N-n+1)\lfloor p;$$

where $\lceil p$ and $\lfloor p$ are the binomial confidence bounds for the same α, c and n, looked up in tables or obtained from Table 5.

For given α, N, n and c, an exact integer solution of (6.1) or (6.2) will usually not exist. Moreover, (6.3) and (6.4) will produce fractional values. We shall not discuss the merits of rounding off to the nearest integer "on the safe side", or of any other system.

For footnotes and an example consult section 1.

TABLE 7 : POISSON, BINOMIAL AND HYPERGEOMETRIC QUANTILES

Wanted: the P-*quantile* k, i.e. given $0 < P < 1$ and given the parameter(s) of the distribution, find k such that the probability of k or less events equals P.

Let ζ be the *standard normal* P-quantile, i.e. $\Phi(\zeta) = P$.

$$k \approx \lambda + \zeta\lambda^{\frac{1}{2}} + (\zeta^2-4)/6 - (\zeta^3+2\zeta)/(72\lambda^{\frac{1}{2}}) + + (3\zeta^4+7\zeta^2-16)/(810\lambda) \tag{7.1}$$

in the Poisson case;

$$k \approx np + \zeta\sigma - \tfrac{1}{2} + (q-p)(\zeta^2-1)/6 + - \{\zeta^3(1+2pq) + \zeta(2-14pq)\}/(72\sigma) \tag{7.2}$$

in the binomial case, where $q = 1-p$ and $\sigma = (npq)^{\frac{1}{2}}$;

$$k \approx nr/N + \zeta\tau - \tfrac{1}{2} + (m-n)(s-r)(\zeta^2-1)/(6N^2) - (72\tau N^4)^{-1} \times \times \{\zeta^3(N^4+2mnN^2+2rsN^2-26mnrs) + \zeta(N^4-14mnN^2-14rsN^2+74mnrs)\} \tag{7.3}$$

in the hypergeometric case, where $m = N-n$, $s = N-r$ and $\tau = (mnrs/N^3)^{\frac{1}{2}}$. The last term of (7.1), (7.2) and (7.3) can be omitted unless P is extremely close to 0 or 1 and/or $\lambda^{\frac{1}{2}}$, σ or τ are small, say less than 2.

TABLE 8 : STANDARD NORMAL AND AUXILIARY VALUES

α	.2	.1	.05	.025	.01	.005
ξ	.842	1.282	1.645	1.960	2.326	2.576
$\Phi(\xi)=1-\alpha$	.8	.9	.95	.975	.99	.995
$\psi=(\xi^2+2)/3$	.90	1.21	1.57	1.95	2.47	2.88
$\omega=(7-\xi^2)/18$	.35	.30	.24	.18	.09	.02

1) For right hand tails, use $P[X \geq h] = 1 - P[X \leq h-1]$, entering the Table with $k = h-1$.

2) $\Phi(u) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{u} \exp(-\frac{1}{2}t^2)\,dt$ denotes the standard normal distribution function, and ξ denotes the upper α standard normal quantile defined by $\Phi(\xi) = 1-\alpha$. Tables of Φ are almost universally available. Values of ξ for some customary choices of α are found in Table 8.

3) In most statistical applications, accurate approximation to probabilities between .005 and .05 or between .95 and .995 will be essential. In such cases one should use the lines marked "for tails" throughout. The influence of terms of higher order, discussed in Appendix B, causes a slight asymmetry in the boundary between tail region and middle one.

4) In cases where p is only just "small" enough in the sense of the Table, it may be useful to know that the Poisson approximations tend to be somewhat less accurate for right hand tails (distribution function near 1).

5) When 1-p is "small", the Poisson approximations can be used after interchanging the roles of successes and failures:
P[k or less successes] = 1 - P[n-k-1 or less failures].

6) It is assumed in Table 3 that the hypergeometric parameters satisfy $n \leq r \leq \frac{1}{2}N$, and in Table 6 that $n \leq \frac{1}{2}N$. This is no restriction, as one can use $P[A \leq a] = P[B \geq n-a] = P[C \geq r-a] = P[D \leq N-n-r+a]$ in the 2×2 table

A	B	n
C	D	N-n=m
r	N-r=s	N

.

7) When a $(1-\alpha)$ two-sided confidence interval is desired, combine the two $(1-\frac{1}{2}\alpha)$ one-sided bounds, i.e. enter the Table with α replaced by $\frac{1}{2}\alpha$.

8) For (5.6) and (5.7), the maximum difference between actual and nominal error probability is less than .001 for $n \geq 12$ and less than .0005 for $n \geq 20$.

Example, illustrating the use of Tables 5 and 6. In a sample of size $n = 20$ from a population of size $N = 200$, $c = 10$ defectives were found. What is the lower bound Lr, with 99% confidence, for the number of defectives in the population?

The answer is $Lr = 50$, as this produces a hypergeometric probability (6.1) of .0098. Without hypergeometric tables one should use approximation (6.4), for which we need the binomial confidence bound p for $\alpha = .01$, $n = 20$, $c = 10$.

Let us first use (5.5) to find Lp. Consulting Table 8 we obtain $[9 + 2.47 - 2.326\{9.91 \times 10.91/16.99\}^{\frac{1}{2}}]/23.94 = .234$. As (5.5) might not be accurate enough for $n = 20$ and $Lp = .234$, one could also try (5.7). Inserting there $L\lambda = 4.129$ (exact) or $L\lambda = 4.132$ from (4.6), cf. example given in Table 4, we find in both cases $Lp = .239$, which is also the exact value for (5.1).

Now (6.4) leads to 49.1 and 50.0 when one inserts $Lp = .234$ and $Lp = .239$ respectively. Thus one obtains 49 from (5.5) and (6.4), and the correct answer 50 from any better Lp inserted into (6.4).

APPENDIX A : REFERENCES

A fairly complete survey of published approximations to the three distribution functions is given in [6]. Confidence bounds are discussed in [1], Poisson quantiles and bounds in [7]. Normal and Poisson approximations to the binomial distribution function are studied with a special error criterion in [4] and [10], whereas normal approximations to a whole class of distribution functions are discussed in [9].

The simple square root approximations (3.4) and (3.5) are believed to be new. For any skew case (3.5) is asymptotically twice as accurate (cf. Appendix B) as the classical χ^2 and normal approximations, and (3.4) is nearly always still better for tails. Obvious limit operations transform (3.4) into (2.2) and (1.2), published in [3,8], and (3.5) into (2.3) and (1.3).

From the general and accurate results of [9] we took (2.7), and (1.5) after a small modification. The general idea of choosing the parameter of the approximating distribution dependent on the argument k of the unknown distribution function, e.g. $(2n-k)p/(2-p)$ instead of np as a Poisson parameter, goes back to Wise and Bolshev: [2] gives (2.5), and [11] gives (3.2) and a complicated formula from which we derived (3.3), see [6] for details.

Solving λ from (1.2) one gets (4.3) and (4.4); the Wilson-Hilferty χ^2 approximation similarly [6] leads to (4.5) and (4.6). The bounds (5.2) and (5.3) come from (2.5). Avoiding the use of a special table of corrections depending on α and c, we adapted (5.6) and (5.7) from [1] at the price of a very slight loss of accuracy. (6.3) and (6.4) use (3.2). The quantile expansions (7.1) and (7.2) are well known.

The remaining recommended formulae are believed to be new.

We have given a few numerical examples and error bounds, but it is virtually impossible to give complete condensed information on errors when so many parameters are involved. We challenge the reader to check a few cases for himself, in order to illustrate the general superiority of the recommended formulae compared to the classical approximations.

APPENDIX B : ASYMPTOTIC EXPANSIONS

The asymptotic theory leading to the recommendations is fully discussed in [5,6,7]. A simple case will be presented here in order to illustrate the train of thought.

For each positive λ and each non-negative integer k there is one fixed value P of the Poisson distribution function, and therefore one unique "exact normal deviate" $\zeta = \zeta(k,\lambda)$ such that $\Phi(\zeta) = P$, i.e.

$$(1) \qquad \Phi(\zeta) = \int_{-\infty}^{\zeta} \phi(t)dt = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\zeta} \exp(-\tfrac{1}{2}t^2)dt = P = \sum_{j=0}^{k} e^{-\lambda}\lambda^j/j!$$

From this transcendental equation an explicit solution for ζ in terms of k and λ is not possible. However, from a Cornish-Fisher type expansion one finds, for $\lambda \to \infty$ and P bounded away from 0 and 1, that:

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

(3) $\Phi\{(k+\frac{1}{2}-\lambda)\lambda^{-\frac{1}{2}}\} = \Phi(\zeta) + \lambda^{-\frac{1}{2}}\phi(\zeta)(\zeta^2-1)/6 + O(\lambda^{-1})$,

(4) $\Phi\{2(k+1)^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}\} = \Phi(\zeta) + \lambda^{-\frac{1}{2}}\phi(\zeta)(4-\zeta^2)/12 + O(\lambda^{-1})$,

(5) $\Phi\{2(k+\frac{3}{4})^{\frac{1}{2}} - 2\lambda^{\frac{1}{2}}\} = \Phi(\zeta) + \lambda^{-\frac{1}{2}}\phi(\zeta)(1-\zeta^2)/12 + O(\lambda^{-1})$.

Let λ be large enough to make the term of order λ^{-1} small compared to the term of order $\lambda^{-\frac{1}{2}}$. Then (2) and (4) achieve a very small error for the special case $\zeta^2 = 4$, i.e. $P = \Phi(\pm 2) = .977$ or .023, and the error of (3) and (5) becomes very small for $\zeta^2 = 1$, i.e. $P = \Phi(\pm 1) = .84$ or .16. Moreover, for any other values of P the error of (2) is minus twice the error of (4), and similarly for (3) and (5), provided that the term $O(\lambda^{-1})$ is negligibly small. This would lead us to prefer (4) for tails, and (5) for the middle part of the distribution, to the classical (2) and (3).

It remains to show that the actual errors of the approximations for moderate and large values of λ follow this asymptotic pattern. The following table of relative errors as percentage of the true value may serve to illustrate that for $\lambda = 30$ the terms of higher order have indeed a modest influence, and dominate only when ζ is close to $\pm$ 1 or $\pm$ 2.

Poisson probability	(2)	(4)	(3)	(5)
$P(X\leq 17) = .0073$	+21	-7	+55	-21
$P(X\leq 19) = .0219$	+2	+2	+26	-11
$P(X\leq 21) = .0544$	-8	+6	+11	-5
$P(X\leq 23) = .1146$	-12	+8	+3	-1
$P(X\leq 25) = .2084$	-13	+8	-1	+1
$P(X\leq 27) = .3329$	-12	+7	-3	+1
$P(X\leq 29) = .4757$	-10	+5	-3	+1

Neither the 1: (-2) ratio nor the optimum at P = .023 and .16 are completely realized, but the agreement is close enough to justify our preference for (4) and (5).

This idea of numerical verification of conclusions derived from asymptotic expansions was also applied to the other approximations mentioned in this paper. Generally speaking the agreement was good, but some exceptions were found, notably for the hypergeometric distribution.

ACKNOWLEDGEMENTS

Many thanks are due to Professors Hemelrijk and Van Zwet, for encouragement and criticism, to Mr. Vehmeyer, for computational assistance, and to the Mathematisch Centrum and Pennsylvania State University for allowing computer use.

REFERENCES

[1] ANDERSON, T.W., and H. BURSTEIN, Approximating the upper binomial confidence limit, J. Amer. Stat. Assoc. 62 (1967), 857-862; Approximating the lower binomial confidence limit, J. Amer. Stat. Assoc. 63 (1968), 1413-1416.

[2] BOLSHEV, L.N., B.V. GLADKOV and M.V. SHCHEGLOVA, Tables for the calculation of B- and Z-distribution functions, Theory of Prob. and its Applic. 6 (1961), 410-419 (English translation of Teoryia Veroyatnostei i ee Primeneniya).

[3] FREEMAN, M.F., and J.W. TUKEY, Transformations related to the angular and the square root, Ann. Math. Stat. 21 (1950), 607-611.

[4] GEBHARDT, F., Some numerical comparisons of several approximations to the binomial distribution, J. Amer. Stat. Assoc. 64 (1969), 1638-1646.

[5] MOLENAAR, W., How to poison Poisson (when approximating binomial tails), Stat. Neerlandica 23 (1969), 19-40 and 241.

[6] MOLENAAR, W., Approximations to the Poisson, binomial and hypergeometric distribution functions, M.C. Tract No. 31, Mathematisch Centrum, Amsterdam 1970.

[7] MOLENAAR, W., Normal approximations to the Poisson distribution, Random Counts in Scientific Work vol. 2, 237-254, edited by G.P. Patil, Pennsylvania State University Press, 1970.

[8] MOSTELLER, F., and J.W. TUKEY, The uses and usefulness of binomial probability paper, J. Amer. Stat. Assoc. 44 (1949), 174-212.

[9] PEIZER, D.B., and J.W. PRATT, A normal approximation for binomial, F, beta and other common, related tail probabilities, J. Amer. Stat. Assoc. 63 (1968), 1417-1456 (part I) and 1457-1483 (part II, by J.W. Pratt alone).

[10] RAFF, M.S., On approximating the point binomial, J. Amer. Stat. Assoc. 51 (1956), 293-303.

[11] WISE, M.E., A quickly convergent expansion for cumulative hypergeometric probabilities, direct and inverse, Biometrika 41 (1954), 317-329.