Approximations of the binomial distribution.

West, Sandra
Boston University

http://hdl.handle.net/2144/14227

Boston University
BOSTON UNIVERSITY
GRADUATE SCHOOL

Thesis
APPROXIMATIONS OF THE BINOMIAL DISTRIBUTION
by

Sandra West
(B.S. Simmons College, 1961)
Submitted in partial fulfillment of the requirements for the degree of

Master of Arts
1963
Approved
by

First Reader ....... [Signature]
Professor of Mathematics

Second Reader ....... [Signature]
Professor of Mathematics
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td></td>
<td>i.</td>
</tr>
<tr>
<td>1 NORMAL APPROXIMATIONS</td>
<td></td>
<td>1.</td>
</tr>
<tr>
<td>1. SIMPLER NORMAL APPROXIMATIONS</td>
<td></td>
<td>1.</td>
</tr>
<tr>
<td>11) VARIANCE STABILIZING TRANSFORMATIONS</td>
<td></td>
<td>14.</td>
</tr>
<tr>
<td>2 POISSON APPROXIMATIONS</td>
<td></td>
<td>20.</td>
</tr>
<tr>
<td>3 GRAM-CHARLIER APPROXIMATIONS</td>
<td></td>
<td>31.</td>
</tr>
<tr>
<td>1) NORMAL GRAM-CHARLIER</td>
<td></td>
<td>31.</td>
</tr>
<tr>
<td>11) POISSON GRAM-CHARLIER</td>
<td></td>
<td>38.</td>
</tr>
<tr>
<td>4 CAMP-PAULSON APPROXIMATION</td>
<td></td>
<td>42.</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td></td>
<td>1.</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>ii.</td>
</tr>
</tbody>
</table>
INTRODUCTION

This thesis is concerned with a discussion of various approximations to the cumulative binomial distribution. A variety of approximating functions have been studied; partly because many times in applications the work required to compute exact probabilities is impractical, and partly for their value in theoretical work.

These approximating functions are useful when specialized tables are not available or do not cover the necessary range. However in order to use these functions advantageously, it is necessary to know something about the conditions under which they may be used. In this paper we will derive various approximations and compare their respective merits.

The approximations studied include, various approximations based on transformations of a binomial distribution into a normal distribution. In particular we discuss the simple normal approximation which is based on the DeMoivre-Laplace Limit Theorem, and the arcsine approximation which is based on the variance stabilizing angular transformation. Also included are the Poisson, and the Poisson modified approximations. Next we discuss the Gram-Charlier approximations; the Normal
Gram-Charlier approximation, which consists of the normal approximation plus an adjustment term matching the skewness of the point binomial, and the Poisson Gram-Charlier approximation. The concluding approximation studied, is the Camp-Paulson, which is based on an entirely different principle from the previous ones; it does not try to match or stabilize the moments of a binomial distribution, but proceeds from the equivalence of a cumulative binomial probability to a probability integral of the variance ratio $F$. 
Consider the situation wherein an experiment is to be carried out independently a certain number, \( n \), of times and at each trial there are only two possible outcomes, \( S \) and \( F \). Attached to the outcome \( S \) there is a probability \( p \), and to \( F \), a probability \( 1-p=q \). We shall assume that these probabilities remain constant throughout the repetitions of the experiment. Such trials are known as Bernoulli trials. When the outcome \( S \), which is usually referred to as a success, is given the probability \( p \), then an acceptable assignment of probability is determined for every choice of the number \( p \) provided only that \( 0\leq p \leq 1 \). Since the outcome \( F \), which is usually referred to as a failure, has as its probability \( q \) which is equal to \( 1-p \), once \( p \) is determined \( q \) is automatically determined.

The sample space of \( n \) Bernoulli trials contains \( 2^n \) pts.; each pt. representing one possible outcome of the compound experiment. We shall be interested in the total number of \( S \)'s appearing in the \( n \) trials, without regard to the order in which the \( S \)'s appear. The probability of exactly \( k \) \( S \)'s and \( n-k \) \( F \)'s appearing in \( n \) trials is given by the following:

\[
b(k; n, p) = \binom{n}{k} p^k q^{n-k}, \quad \text{where } k=0,1,2,\ldots,n.
\]

We introduce the random variables \( X_1, X_2, \ldots, X_n \) defined as:

---

Parzen, Pg. 102
If S appears at the kth trial
\( X_k = 1 \)

If F appears at the kth trial
\( X_k = 0 \)

We may then introduce the variable
\( S_n = X_1 + X_2 + \ldots + X_n \)

which equals the number of S's appearing in n independent trials. Here we have a sum of n identically distributed random variables; that is, each \( X_k \) is distributed according to the same Bernoulli probability law:

\[
P(X_k=1) = p
\]

\[
P(X_k=0) = q \quad \text{for } k=1,2,3,4,\ldots,n
\]

Then

\[
P\{S_n = k\} = b(k;n,p).
\]

\( S_n \) is a random variable and \( b(k;n,p) \) is the so-called Binomial distribution.

It is useful for future reference to derive the mean and 2nd and 3rd central moments for the binomial distribution.

The expected value is defined by the equation:

\[
E(X) = \sum_{k=1}^{n} k b(k;n,p)
\]

The expected value is identical with the population mean, \( \mu \).

The 2nd moment about the mean is given by the equation:

\[
\sigma^2_X = Var(X) = E(X^2) - \mu^2 = E(X^2) - \mu^2
\]

For the binomial distribution, \( X_k \) can take the values 0 or 1 with probabilities \( q \) and \( p \) respectively. Thus:

\[
E(X_k) = 1 \cdot p + 0 \cdot q = p
\]

\[
E(X_k^2) = 1 \cdot p + 0 \cdot q = p
\]

\[
\therefore Var(X_k) = p - p^2 = p(1-p) = p \cdot q
\]

Thus for \( S_n = X_1 + X_2 + X_3 + \ldots + X_n \) where the \( X \)'s are independent and identically distributed,

\[
E(S_n) = E(X_1 + X_2 + X_3 + \ldots + X_n) = EX_1 + EX_2 + EX_3 + \ldots + EX_n = np
\]
\begin{align*}
\text{Var}(S_n) &= \text{Var}(X_1 + X_2 + X_3 + \ldots + X_n) = n \text{Var}(X) = npq \\
\text{If } \bar{X} \text{ is defined as } \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{S_n}{n},
\end{align*}

then \( E(\bar{X}) = \frac{E(S_n)}{n} = p \) and \( \text{Var}(\bar{X}) = \frac{\text{Var}(S_n)}{n^2} = pq \).

The third moment about the mean is given by the equation:

\[ M_3 = E(X-\mu)^3 \]

For \( S_n \) the 3rd moment about the mean (the central moment) is:

\[ E(S_n-np)^3 = E(X_1 + X_2 + \ldots + X_n-np)^3 = E\left[ (X_i-np)^3 (X_2-p) \ldots (X_n-np) \right]^3 \]

where \( i\neq j \neq k \)

\[ Ez = Ex - p = p - p = 0 \]

\[ (EZ)^3 = Ez_{i}Z_{j}Z_{k} \text{ where } i\neq j \neq k \]

\[ Ez_{i}Z_{j}Z_{k} = Ez_{i}Ez_{j}Ez_{k} \]

\[ Ez_{i}^2Z_{j} = Ez_{i}^2Z_{j} = 0 \]

\[ Ez_{i}^3 = Ez_{i}(x_{i}^3 - 3x_{i}^2p + 3x_{i}p^2 - p^3) = Ez_{i}^3 - 3px_{i}^2 + 3p^2x_{i} - p^3 = p - 3p^2 + 3p^3 - p^3 \]

\[ \sum_{i=1}^{n} Ez_{i}^3 = n(p^3 - 2p^2 + 2p^3) = np(1 - 2p)(1 - p) = npq(1 - p - p) = npq(q - p) \]

\[ \text{E}(S_n-np)^3 = npq(q - p) \]

The binomial distribution has a simple form, but this does not mean that the numerical values are easily computed. In many applications the labor required to compute exact probabilities is prohibitive. To help this situation tables have been published and a variety of approximating functions have been studied. In order to use these functions it is necessary to know something about the conditions for which one approximation may be used as opposed to another. In this
paper we shall see various approximations to the cumulative binomial probability which is defined as:

\[ P(X \leq k) = B(k; n, p) = \sum_{i=0}^{k} \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \]

\( k \geq 0, k \) has integral values from 0 to \( n \)

Consider:

\[ B(n-k-1; n, q_\ast) = \sum_{i=0}^{n-k-1} \frac{n!}{i!(n-i)!} q_\ast^i (1-q_\ast)^{n-i} \]

\[ = p^n + \frac{n!}{1!(n-1)!} (1-p) + \ldots \frac{n!}{(n-k)!} (1-p)^{n-k} \]

\[ B(k; n, p) = (1-p)^n + \frac{n!}{1!(n-1)!} (1-p)^{n-1} + \ldots \frac{n!}{k!(n-k)!} (1-p)^{n-k} \]

\[ 1 - B(k; n, p) = \frac{n!}{(k+1)!(n-k-1)!} (1-p)^{n-k} + \ldots \frac{n!}{n!(1-p)^n} - B(n-k-1; n, q_\ast) \]

Thus this probability has the symmetric property; therefore when comparing approximations of \( B(k; n, p) \) for various values of \( p \), we only have to consider values of \( p \) smaller than \( \frac{1}{2} \).

We are often interested in the probability that the number of successes lies between preassigned limits, \( \alpha \) and \( \beta \). If \( \alpha \) and \( \beta \) are integers (\( \alpha, \beta \)), then we define the event as \( \alpha \leq S_n \leq \beta \). It's probability is:

\[ P(\alpha \leq S_n \leq \beta) = b(\alpha; n, p) + b(\beta; n, p) + \ldots b(\beta; n, p) = B(\beta) - B(\alpha - 1) \]

This sum may involve many terms and a direct evaluation may be impractical. Fortunately, whenever \( n \) is large, the normal distribution function can be used to derive a simple approx. to this probability. This discovery is due to DeMoivre and Laplace. It is one of the most used approximations, however its importance goes far beyond the domain of numerical calculations.
It will be convenient to introduce the following terminology. Generally, if the distribution of a random variable $X$ depends on a parameter $n$, and if two quantities $\mu_n$ and $\sigma_n$ can be found such that the distribution function of the variable

$$\frac{X - \mu_n}{\sigma_n} \text{ tends to } \frac{t}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

as $n \to \infty$, we shall say that $X$ is asymptotically normal $(\mu_n, \sigma_n)$. This does not imply that the mean and standard deviation of $X$ tend to $\mu_n$ and $\sigma_n$, nor even that these moments exist, but is simply equivalent to saying that we have for any interval $(\alpha, \beta)$ not depending on $n$

$$\lim_{n \to \infty} P(\alpha < X - \mu_n < \beta) = \frac{1}{\sqrt{2\pi n}} \int_{\alpha - \beta}^{\beta} e^{-\frac{t^2}{2}} \, dt$$

The so-called Central Limit Theorem may now be expressed in the following way:

Whatever be the distribution of the independent variable $X$, subject to certain general conditions, the sum $X = X_1 + X_2 + \ldots + X_n$ is asymptotically normal with mean $\mu$, and variance $\sigma^2$. If in this theorem, we let $X$ be the sum of $n$ Bernoulli variables we get the DeMoivre-Laplace Limit Theorem. That is

$$P(\alpha < X < \beta) \equiv \frac{1}{\sqrt{2\pi npq}} \int_{\alpha - \beta}^{\beta} e^{-\frac{1}{2} \left( \frac{X - np}{\sqrt{npq}} \right)^2} \, dx$$

A slightly better approximation has been found to be

$$P(\alpha < X < \beta) \equiv \frac{1}{\sqrt{2\pi npq}} \int_{\alpha - \beta}^{\beta + \frac{1}{2}} e^{-\frac{1}{2} \left( \frac{X - np}{\sqrt{npq}} \right)^2} \, dx$$

\[\text{Cramer, Pg. 214}\]

\[\text{Cramer, Pg. 198}\]
This adjustment of 1/2 unit is to take into account the adapting of the normal distribution, which is continuous to the binomial distribution which is discrete.

When we use an approximating formula instead of an exact one, there is always the question to consider: How large is the committed error? If, as is usually done, the question is left unanswered, the derivation of Laplace's formula is relatively easy. However to estimate the error comparatively long and detailed investigation is required.

According to Uspensky, the probability that \( \left| \frac{S_n - np}{\sqrt{npq}} \right| < \frac{4}{\sqrt{2}} \) is given by:

\[
P = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} \, du + \frac{(1 - \theta_i - \theta_j)}{\sqrt{2\pi npq}} \quad \text{and where} \quad \frac{1}{\sqrt{npq}} < \frac{20 + 25|p-q|}{npq} + e^{-\frac{3}{2}npq}
\]

and where \( \theta_i \) and \( \theta_j \) are the respective fractional parts of \( np + \sqrt{npq} \) and \( nq - \sqrt{npq} \) respectively.

Referring to the above expression for the probability of the inequalities \( t_1 \sqrt{npq} < S_n - np < t_2 \sqrt{npq} \) and supposing that the number of trials \( n \), increases indefinitely while \( t_1 \) and \( t_2 \) remain fixed, we see Laplace's Limit Theorem.

\[
P \left\{ \frac{t_1}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{t_2}{\sqrt{npq}} \right\} \Rightarrow \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{1}{2}u^2} \, du \quad \text{as} \ n \to \infty
\]

To form an idea of the accuracy to be expected by using the foregoing approximating formulas, let us take up a numerical example.

Uspensky, Pg. 130
Let \( n = 200 \), \( p = q = \frac{1}{2} \), \( 95 \leq S_n \leq 105 \). 

The exact expression of the probability that \( S_n \) will satisfy the above inequality is:

\[
P = 2 \left[ \frac{200!}{95!105!} \left( \frac{1}{2} \right)^{200} \right] + \frac{200!}{96!104!} \left( \frac{1}{2} \right)^{200} + \frac{200!}{97!103!} \left( \frac{1}{2} \right)^{200} + \ldots + \frac{200!}{100!100!} \left( \frac{1}{2} \right)^{200}
\]

\[
P = \frac{200!}{100!100!} \left[ 1 + 2 \left( \frac{100}{101} + \frac{100.99}{102} + \frac{100.99.98}{103} + \frac{100.99.98.97}{104} + \ldots + \frac{100.99.98.97.96}{105} \right) \right]
\]

\( P \) is found to be \( .56325 \) and this may be regarded as correct to 5 decimal places.

Using the approximate formula:

\[
t = \sqrt{\frac{npq}{t}} = \sqrt{\frac{50}{5}} = 5 \quad t = \frac{1}{\sqrt{2}} = .707107
\]

\[
2 \int_0^{\frac{T}{\sqrt{2}\pi}} e^{-\frac{1}{2}v^2} dv = .52050 \quad \text{and} \quad e^{-\frac{1}{2}T^2} = .04394
\]

Thus \( P = \frac{2}{V_{2\pi}} \int_0^{\frac{T}{\sqrt{2}\pi}} e^{-\frac{1}{2}v^2} dv + \left( 1 - \theta - \theta \right) e^{-\frac{1}{2}T^2} = .52050 + .04394 = .56444\)

This is greater than the true value of \( P \) by \( .00119 \).

The theoretical limit of the error: \( |M| < \frac{.20 + .25}{25} \left| \frac{1}{2} - \frac{1}{2} \right| = \frac{1}{25} = .004 \)

So that actually, using the above formula gives an even closer approximation than can be expected theoretically.

When \( npq \) is large the 2nd term in Laplace's formula ordinarily is omitted and the probability is computed by using a simpler expression:

\[
P = \frac{2}{\sqrt{2\pi}} \int_0^{\frac{T}{\sqrt{2}\pi}} e^{-\frac{1}{2}v^2} dv
\]

In our case this expression would give \( P = .52050 \).
Here we have an error about .043 which is about 8% of the exact number. Such a comparatively large error is explained by the fact that in our example npq=50 is not large enough.

There is more than one way of judging the closeness of a given approximation. Uspensky in discussing the normal approximation considers errors in the cumulative binomial probability. Others measure their errors in terms of the normal deviate corresponding to the cumulative binomial prob.; this method would reduce the differences near the center of the distribution, and magnify those in the tails. Still other criteria are possible, such as the relative error in the cumulative probability. Different criteria serve different purposes, and any choice seems to be somewhat arbitrary. Table 1.1 gives the maximum error of the normal approximation $\Phi(X)$ to the binomial distribution. Where $\Phi(X)$ is a normal distribution having the same mean and variance as the binomial distribution.

If we let

$$\Phi(Z) = B_1(k;n,p) = \int_0^\infty e^{-\frac{t^2}{2}} \cdot dt$$

and if we let:

$$B(k;n,p,) = \sum_{i=0}^{k} \frac{n^i}{i!(n-1)!} p^i (1-p)^{n-i}$$

if $k > 0$

The closeness of the approximation $B_1(k;n,p)$ was judged by examining the maximum error which is defined as follows:

$$E_1(k;n,p) = B_1(k;n,p) - B(k;n,p)$$
The maximum error
\[ M_i(np) = \max_{j,k} \left| E_i(j;n,p) - E_i(k;n,p) \right| \]
where \( j \) and \( k \) can take any integral values from \( 0 \) to \( n \).
That is, the maximum error is defined as the largest possible error which can arise in estimating any sum of consecutive binomial terms with the specified parameters.
Table 1.1 lists the maximum errors of \( B_i \) as a function of \( n \) and \( p \) and also as a function of \( n \) and \( np \). For constant \( n \) the maximum error \( M_i(np) \) decreases as \( p \) increases to \( \frac{1}{2} \) (except when \( p \) is very near 0, where the trend is reversed.) For constant \( p \) it decreases with increasing \( n \). When the mean \( np \) is held constant , the maximum error increases with increasing \( n \) up to a limiting value which represents the error in the normal approximation to the Poisson distribution. (This we will discuss in the next chapter.) According to Raff \( M_i(np) \) is always less than \( 0.140/\sqrt{npq} \).
<table>
<thead>
<tr>
<th>values of $p$</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>values of $n$</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td>.126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td>.125</td>
<td>.082</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.008</td>
<td></td>
<td></td>
<td></td>
<td>.082</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.010</td>
<td></td>
<td></td>
<td></td>
<td>.124</td>
<td>.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.020</td>
<td></td>
<td></td>
<td></td>
<td>.122</td>
<td>.080</td>
<td>.046</td>
<td>.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.040</td>
<td></td>
<td></td>
<td></td>
<td>.118</td>
<td>.077</td>
<td>.030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.080</td>
<td></td>
<td></td>
<td></td>
<td>.071</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.100</td>
<td></td>
<td></td>
<td></td>
<td>.158</td>
<td>.106</td>
<td>.060</td>
<td>.040</td>
<td>.027</td>
<td></td>
</tr>
<tr>
<td>.200</td>
<td></td>
<td></td>
<td></td>
<td>.086</td>
<td>.054</td>
<td>.032</td>
<td>.022</td>
<td>.015</td>
<td></td>
</tr>
<tr>
<td>.300</td>
<td></td>
<td></td>
<td></td>
<td>.054</td>
<td>.032</td>
<td>.019</td>
<td>.013</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>.400</td>
<td></td>
<td></td>
<td></td>
<td>.024</td>
<td>.016</td>
<td>.009</td>
<td>.006</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>.500</td>
<td></td>
<td></td>
<td></td>
<td>.011</td>
<td>.005</td>
<td>.002</td>
<td>.001</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

values of $np$

<p>| 0.000 | .45 | .186 | .50 | .158 | .185 |
| 1.000 | .086 | .106 | .118 | .122 | .124 | .125 | .126 | .126 |
| 1.500 | .054 | .126 | .109 | .082 | .082 | .082 | .083 |
| 2.000 | .024 | .054 | .071 | .077 | .080 | .082 | .082 | .083 |
| 2.500 | .011 | .060 | .073 |</p>
<table>
<thead>
<tr>
<th>values of np</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.00</td>
<td>.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.00</td>
<td>.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.00</td>
<td>.005</td>
<td>.032</td>
<td>.040</td>
<td>.044</td>
<td>.046</td>
<td>.046</td>
<td>.047</td>
<td></td>
</tr>
<tr>
<td>7.50</td>
<td></td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.00</td>
<td>.009</td>
<td>.022</td>
<td>.027</td>
<td>.030</td>
<td>.032</td>
<td>.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.00</td>
<td></td>
<td>.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.00</td>
<td>.006</td>
<td>.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.023</td>
</tr>
<tr>
<td>25.00</td>
<td></td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.00</td>
<td></td>
<td>.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.018</td>
</tr>
<tr>
<td>40.00</td>
<td></td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.016</td>
</tr>
<tr>
<td>50.00</td>
<td></td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.014</td>
</tr>
</tbody>
</table>
As an application of the simple normal approximation let us find confidence intervals for the parameter \( p \), that appears in the binomial distribution.

Consider the following test:

\[ H_0 \quad p = p_0 \]
\[ H_1 \quad p \neq p_0 \]

the critical statistic is \( \hat{p} = \frac{\bar{x}}{n} \)
and the critical region is: accept \( H_0 \) if
\[
|Z| = \frac{|\hat{p} - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} < Z_{1-\alpha/2}
\]
where \( P(Z > Z_{1-\alpha/2}) = \frac{\alpha}{2} \)
and \( Z \) is approximately normal with mean=0, and variance =1.

Thus, given an observed value of \( p \) all hypothetical values of \( p_0 \) which will be accepted are those which satisfy the above inequality. Let's call \( Z_{1-\alpha/2} = r \), and square both sides of the inequality:
\[
\frac{(\hat{p} - p_0)^2}{p_0(1-p_0)} < r^2 \quad \text{and} \quad p_0(1-p_0) > 0
\]

\[
(\hat{p} - p_0)^2 n - r^2 p_0 (1-p_0) < 0
\]
Let's consider this inequality as a function of \( p_0 \) that is:
\[
f(p_0) = (\hat{p} - p_0)^2 n - r^2 p_0 (1-p_0)
\]
The graph of \( f(p_0) \) against \( p_0 \) is a parabola opening upwards, since the coefficient of \( p_0 \) is positive.
\[
f(p_0) = p_0^2 (n + r^2) - p_0(2np + r^2) + n\hat{p}^2
\]
The value of \( p_0 \) which satisfies the inequality \( f(p_0) < 0 \)
lies between the two roots of the quadratic equation

\[ f(p_o) = 0 \]

Solving this equation for \( p_o \), using the quadratic formula, we get:

\[ \eta_{p_o} = \frac{2np + r^2 + r\sqrt{r^2 + 4np(1 - \hat{p})}}{2(n + r^2)} \]

Since any value of \( p_0 \) lying between these two roots will satisfy \( f(p_o) < 0 \), the two values of \( p_o \) given by the above equation are the confidence limits for \( p_o \).

If \( r \) is small compared to \( n \), we get:

\[ p_o = \frac{2n\hat{p} + r\sqrt{4n\hat{p}\hat{q}}}{2n} = \hat{p} + \frac{r}{\sqrt{n}} \sqrt{n\hat{p}\hat{q}} \]

\[ = \bar{x} + \frac{r}{\sqrt{n}} \sqrt{\frac{c^2}{n}} \]

with \( c^2 \) estimated by \( \frac{\hat{c}^2}{\hat{n}} = \bar{x}(1 - \bar{x}) \)

---

Alexander, Pg. 202
VARIANCE STABILIZING TRANSFORMATIONS

A different type of approximation is based on the so-called variance stabilizing distribution.
Suppose that a relation

$$\sigma_\chi = h(\mu_\chi)$$

between the mean and standard deviation of the variable \(X\) exists.
We want to find \(Y = g(X)\) such that \(\sigma_\chi^2 = \text{constant}\)
From experience, it has been found that such transformations, usually transform a distribution into a distribution that is more nearly normal.

By Taylor's formula we can write:

$$g(X) \approx g(\mu) + (X-\mu) \cdot g'(\mu)$$

so that

$$\sigma_\chi^2 \{g(X)\} \approx \sigma_\chi^2 \{X\} \cdot g'(\mu_\chi)$$

since \(\sigma_\chi = h(\mu_\chi)\)

$$\sigma_\chi^2 \{g(X)\} \approx h(\mu_\chi) \cdot g'(\mu_\chi)$$

If we choose the transformation function, \(g(X)\), in such a manner that the standard deviation \(\sigma_\chi^2 \{g(X)\}\) is the same for all distributions, we have:

$$h(\mu_\chi) \cdot g'(\mu_\chi) = c$$

where \(c\) is a constant, or

$$g'(\mu_\chi) = c/h(\mu_\chi)$$

Thus, the transformation function is determined by the equation

$$g(X) = c \int \frac{dX}{h(X)}$$

Hald, Pg. 176
Let us consider $X = X/n$, where $X$ is a binomial variable. On the basis of the relation between the mean, $p$, and the variance, $p(1-p)/n$ for $X$, we may determine a function $Y = g(X)$ in such a manner that the variance of the transformed variable $Y$ is independant of $p$.

We have

$$\sigma = \sqrt{p(1-p)/n}$$

where in the previous discussion the $h$ is now replaced by $p$, and $h(X) = \frac{\sqrt{X(1-X)}}{n}$ and

$$g(X) = c \int \frac{dX}{h(X)} = \sqrt{\frac{n}{\pi}} \int \frac{dX}{\sqrt{X-X^2}}$$

Integrating we get:

$$g(X) = \sqrt{n} \cdot c \cdot 2 \cdot \arcsin \sqrt{X} = 2 \cdot \arcsin \sqrt{X}$$  where $c = \frac{1}{\sqrt{n}}$

Previously we found that

$$\sigma^2 \{ g(X) \} \approx h(p) \cdot g'(p)$$

thus

$$\sigma^2 \{ g(X) \} \approx h(p) \cdot g'(p) = \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{n} \cdot c \cdot \frac{1}{\sqrt{p(1-p)}} = c = \frac{1}{\sqrt{n}}$$

Thus

$$g(X) - g(X) = 2 \left[ \arcsin \sqrt{X} - \arcsin \sqrt{p} \right] / \sqrt{n}$$

We can write

$$u = (2 \cdot \arcsin \sqrt{X} - 2 \cdot \arcsin \sqrt{p}) / \sqrt{n}$$

is approximately normally distributed with parameters $(0,1)$.

In analogy with our previous approximation

$$P \{ X \} \approx \Phi \left( \frac{X+\frac{n-p}{np(1-p)}}{\sqrt{n}} \right)$$

we have

$$P \{ X \} \approx \Phi \left( \frac{2 \cdot \arcsin \sqrt{X} + \frac{1}{n} - 2 \cdot \arcsin \sqrt{p}}{\sqrt{n}} \right)$$

$^9$ Hald, Pg. 685
Observations have indicated that the preceding two approximation formulas, usually lead to deviations from the exact values of the same order of magnitude, but with opposite signs. So that the mean of the two approximations usually gives a considerably better approximation to the exact value than either of the two single values. However, this new approximation has only been applied to a certain range. That is, an approximation to \( P\{X \geq x \} \) for \( 0.05 < P < 0.05 \) and \( 0.95 < P < 0.995 \) we may apply \( \Phi(u) \) for

\[
u = \frac{1}{2}(u_1 + u_2)
\]

where

\[
u_1 = \left( x + \frac{1}{2n} - \frac{\theta}{\sqrt{n}} \right)
\]

and

\[
u_2 = (2 \cdot \arcsin \sqrt{x + \frac{1}{2n}} - 2 \cdot \arcsin \sqrt{p}) \sqrt{n}
\]

Table 1.2 gives the maximum error \( M_2(np) \) occurring when we use the approximation:

\[
B_2(k; n, p) = \left[ 2 \sqrt{n} \cdot (\arcsin \sqrt{k + \frac{1}{n}} - \arcsin \sqrt{p}) \right]
\]

The maximum errors \( M_2(np) \) follow a pattern similar to those of the normal approximation, except that the increasing trend with decreasing \( p \) goes all the way to \( p=0 \).

Hald, Pg. 686

AMS. Vol. 44 Pg. 174-212
Table 1.2

<table>
<thead>
<tr>
<th>values of ( p )</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.166</td>
<td>0.098</td>
<td>0.043</td>
<td>0.028</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.084</td>
<td>0.043</td>
<td>0.023</td>
<td>0.015</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>0.048</td>
<td>0.024</td>
<td>0.013</td>
<td>0.009</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>0.030</td>
<td>0.015</td>
<td>0.007</td>
<td>0.004</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.024</td>
<td>0.012</td>
<td>0.004</td>
<td>0.002</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>values of ( np )</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.579</td>
</tr>
<tr>
<td>0.05</td>
<td>0.166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.185</td>
</tr>
<tr>
<td>0.10</td>
<td>0.084</td>
<td>0.098</td>
<td></td>
<td></td>
<td></td>
<td>0.112</td>
</tr>
<tr>
<td>0.15</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.080</td>
</tr>
<tr>
<td>0.20</td>
<td>0.030</td>
<td>0.043</td>
<td>0.056</td>
<td></td>
<td></td>
<td>0.059</td>
</tr>
<tr>
<td>0.25</td>
<td>0.024</td>
<td></td>
<td>0.042</td>
<td></td>
<td></td>
<td>0.050</td>
</tr>
<tr>
<td>0.30</td>
<td></td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
<td>0.044</td>
</tr>
<tr>
<td>0.35</td>
<td></td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td>0.037</td>
</tr>
<tr>
<td>0.40</td>
<td></td>
<td></td>
<td>0.023</td>
<td>0.028</td>
<td></td>
<td>0.032</td>
</tr>
<tr>
<td>0.45</td>
<td></td>
<td></td>
<td>0.013</td>
<td></td>
<td></td>
<td>0.026</td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td>0.019</td>
<td>0.019</td>
<td>0.022</td>
</tr>
<tr>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td>0.015</td>
<td></td>
<td>0.015</td>
</tr>
<tr>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td>0.010</td>
<td></td>
<td>0.015</td>
</tr>
<tr>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td>0.006</td>
<td></td>
<td>0.012</td>
</tr>
<tr>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
<td></td>
<td>0.010</td>
</tr>
</tbody>
</table>

"JASA Vol. 51 pg. 298"
The use of the arcsine transformation has been greatly facilitated and many new applications have been developed by F. Mosteller and J. W. Tukey, in "The Uses and Usefulness of Binomial Probability Paper."

Probability paper has the property that the graph of a cumulative normal distribution is a straight line. This is accomplished by using a special ruling to the ordinate axis, namely a scale where instead of the u-values the corresponding values of \( \Phi(u) \) are marked. The basic idea here, of the binomial probability paper, is to plot the sample point \((n-x,x)\) on double-square root paper; that is, in reality we plot \((\sqrt{n-x}, \sqrt{x})\). This point will lie somewhere on a quarter-circle with radius \( \sqrt{n} \). The angle made by the line connecting the origin with the sample point and the horizontal axis is \( \arcsin \sqrt{x} \). Repeated random samples of size \( n \) will give sample points varying at random on the circle, their standard error measured on the circle being \( \frac{1}{\sqrt{n}} \).

We often use such transformations so that we can compare the transformed binomial data with percentage points of the standard normal distribution to make approximate significance test or to set approximate confidence intervals.

Consider

\[
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-\frac{1}{2} t^2} dt = \text{Prob.}\{ X \leq k \mid \text{binomial, } n,p \}.
\]

That is, we would like to use the normal deviate \( K \) exceeded with the same probability as the number of successes \( X \) from \( n \) in a binomial distribution, with \( np \).
Various approximations to $K$ have been given by Tukey.

For example:

$$L^* = L + (L-2) \cdot (L+2) / 12 \cdot \left( \frac{1}{\sqrt{n} \cdot r + t} - \frac{1}{\sqrt{n} \cdot q + t} \right)$$

where $L = \frac{X + \frac{1}{n} - np}{\sqrt{npq}}$

$M^* = M$ modified by a term like that in $L^*$

where $M = 2 \sqrt{n+1} \left( \arcsin \sqrt{\frac{k+1}{n+t}} - \arcsin \sqrt{p} \right)$

in addition

$$N = 2 \left( \sqrt{(k+1) \cdot q} - \sqrt{(n-k) \cdot p} \right)$$

$$N = N + (N + 2p - 1) / 12 \cdot \sqrt{E} \quad E = \text{lesser of } np \text{ and } nq$$

$N^* = N$ modified by a term like that in $L^*$

$$N = N^* + (N^* + 2p - 1) / 12 \cdot \sqrt{E} \quad E = \text{lesser of } np \text{ and } nq$$

Taking an upper limit of 2.5 or 3.5 on $|K|$, and a lower limit of .01, 1 or 4 on $np$, the greatest observed errors of the transformations were smallest for $N^{**}$, $N^*$ and $M^*$ and largest for the direct approximations $L$ and $L^*$. 
2. POISSON APPROXIMATIONS

Another type of approximation is the Poisson approximation. In many applications we deal with Bernoulli trials where, comparatively speaking, $n$ is large and $p$ is small, whereas the product $\lambda = np$ is of moderate magnitude. In these cases it is convenient to use an approximating formula to $b(k; n, p)$ which is due to Poisson, which can be derived as follows:

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$k=0$  $b(0; n, p) = (1 - p)^n \quad \lambda = np \quad p = \frac{\lambda}{n}$

$$= (1 - \frac{\lambda}{n})^n \quad \Rightarrow \quad e^{-\lambda} \quad \text{as} \quad n \to \infty$$

for fixed $k$:  $b(k; n, p) = \binom{n}{k} q^{n-k}$ and $b(k-1; n, p) \approx (k^{-1}) p$. 

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{n!}{(n-k)!} \frac{(k-1)! (n-(k-1))!}{n!} \frac{p^n q^{-k}}{p^{k-1} q^{n-k+1}}$$

$$= \frac{(n-k+1)p}{k} \quad = \frac{\lambda - (k-1)p}{kq} \quad = \frac{\lambda - (k-1)p}{k(q-\frac{\lambda}{n})} \quad \approx \frac{\lambda}{k}$$

$$b(0; n, p) = e^{-\lambda} \quad \Rightarrow \quad b(1; n, p) = \lambda e^{-\lambda}$$

Assume that for $k=L$

$$b(L; n, p) = \frac{\lambda e^{-\lambda}}{L!}$$

Would like to prove for $k=L+1$:  $b(L+1; n, p) \approx \frac{e^{-\lambda}}{(L+1)!} \lambda^{L+1}$
We assumed that \( b(L;n,p) = \frac{\lambda^L}{L!} \).

\[
\therefore b(L+1;n,p) = \frac{\lambda}{L+1} \frac{\lambda^L e^{-\lambda}}{L!} = \frac{\lambda^{L+1} e^{-\lambda}}{(L+1)!}
\]

Generally we see by induction that:

\[ b(k;n,p) \approx \frac{e^{-\lambda} \lambda^k}{k!} \]

This is the famous Poisson approximation to the binomial distribution. It is customary to call: \( p(k;\lambda) = \frac{e^{-\lambda} \lambda^k}{k!} \)

Thus \( p(k;\lambda) \) is an approximation to \( b(k;n,\frac{\lambda}{n}) \) when \( n \) is sufficiently large. As we have seen the error of the normal approximation will be small if \( npq \) is large. On the other hand if \( n \) is large and \( p \) small the terms \( b(k;n,p) \) will be found to be near the Poisson probability \( p(k;\lambda) \) with \( \lambda = np \).

If \( \lambda \) is small then only the Poisson approximation can be used. However, if \( \lambda \) is large we can use either the normal or the Poisson approximation. This implies that for large values of \( \lambda \) it must be possible to approximate the Poisson distrib. by the normal distribution. This is true, for fixed values of \( \lambda, \beta \) and \( k \), the following difference tends to 0 as \( \lambda \to \infty \):

\[
\frac{e^{-\lambda} \lambda^k}{k!} - \frac{1}{2\pi} \int_{K-\lambda-\frac{1}{2}}^{K-\lambda+\frac{1}{2}} e^{-\frac{y^2}{2}} \, dy \to 0 \quad (\lambda \to \infty)
\]

We may consider this as follows:

We are given the distribution of a sum of independent

\[ Feller, \text{ Pg. 143} \]
identically (Bernoulli) distributed random variables, and we have arrived at two limiting distributions for such a sum by means of two essentially different limiting processes. In the case of the normal approximation to the binomial we hold \( p \) constant and let \( n \to \infty \). For the Poisson approximation to the binomial, we allow both \( n \) and \( p \) to vary, but in such a way that \( n \to \infty \) and \( p \to 0 \) in a manner that leaves \( \lambda \) bounded.

Consider the following configuration:

\[
\begin{align*}
\pi_1 &: X_{11} \\
\pi_2 &: X_{21}X_{22} \\
\pi_3 &: X_{31}X_{32}X_{33} \\
\vdots
\end{align*}
\]

\[
\pi_i &: X_{i1}X_{i2}X_{i3} \cdots X_{in} \\
\vdots
\]

In the first case above we are changing \( n \) alone as we consider successive rows of mutually independent variables; i.e., \( p_i = p \) (\( i = 1, 2, \ldots \)). In the second of the cases we must consider the fact that the \( p \)'s are changing as we move to different rows with increasing \( n \). Actually this configuration is a special case of the more general situation with a double sequence \( X_{n1}, X_{n2}, \ldots X_{nn} \) (\( n = 1, 2, \ldots \)) of random variables, independent for each choice of \( n \), but not necessarily identically distributed.

To get rough error estimates for the Poisson approximations of the binomial distribution, we can proceed as follows:
\[ b(k; n, p) = \frac{n!}{(n-k)k!} p^k (1-p)^{n-k} \quad k \geq 0 \quad \lambda = np \text{ so that } p = \left( \frac{\lambda}{n} \right) \]

\[
\frac{n!}{k!(n-k)!} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} = \frac{n!}{(n-k)!} \left( \frac{\lambda}{k!} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} = b(k; n, p)
\]

To get the lower bound:

\[
\frac{n!}{(n-k)!} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} \geq \frac{n!}{(n-k)!} \left( \frac{\lambda}{k!} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} = b(k; n, p)
\]

Thus \( \frac{\lambda}{k!} \left( 1 - \frac{\lambda}{n} \right)^{n-k} \geq b(k; n, p) \geq \frac{\lambda}{k!} \left( 1 - \frac{k}{n} \right)^{n-k} \left( 1 - \frac{\lambda}{n} \right)^{n-k} \)

By refining the preceding inequality we can show that:

\[
p(k; \lambda) e^{\frac{\lambda k}{n-k}} \geq b(k; n, p) \geq b(k; n, p) e^{-\frac{k^2}{2(n-k)}} \times \frac{\lambda^n}{n!} \cdot \rho(k; \lambda)
\]

Before we can do this we need to derive another inequality.

**By Taylor's expansion:** \( \log \left( \frac{1}{1-t} \right) = t + \frac{1}{2} t^2 + \frac{1}{3} t^3 + \cdots \quad \Rightarrow -1 < t < 1 \)
for $0 < t < 1$; \[t < t + \frac{1}{3} t^2 + \frac{1}{3} t^3 + \cdots < t + \frac{t^2}{1-t} \]

\[t < \log \frac{t}{1-t} < \frac{t}{1-t} \]

multiply by $n'$: \[n' \cdot t < n' \cdot \log \frac{t}{1-t} < n' \cdot \frac{t}{1-t} \]

thus \[e^{n't} < \left( \frac{t}{1-t} \right)^{n'} < e^{n' \frac{t}{1-t}}\]

and \[e^{-\left( \frac{t}{1-t} \right)^{n'}} < e^{-n't} \]

now: \[p(k; \lambda) e^{\frac{k_1}{n}} = e^{-\lambda \frac{1}{k_1!}} e^{\frac{k_1}{n}} = \frac{\lambda}{e^{\frac{1}{k_1!}}} (n-k) \]

according to inequality (A) with $n' = (n-k)$, $t = \lambda$, $0 < \frac{\lambda}{n} \leq 1$

\[e^{-\frac{\lambda}{n}(n-k)} > \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]

Thus \[p(k; \lambda). e^{\frac{k_1}{n}} = \frac{\lambda^{n-k}}{k_1!} e^{-\frac{\lambda}{n} (n-k)} > \frac{\lambda^{n-k}}{k_1!} \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]

\[p(k; \lambda) . e^{-} \geq b(k; n, p) \]

For the lower bound:

\[p(k; \lambda) . e^{- \frac{k^2}{n-k}} e^{- \frac{\lambda^2}{n-k}} = e^{- \lambda / \frac{1}{k_1!}} e^{- \frac{\lambda^2}{n-k}} \]

\[e^{- \lambda / \frac{1}{k_1!}} e^{- \frac{\lambda^2}{n-k}} = e^{- \left( \frac{n-k}{n-k} \right)} = e^{- \left( \frac{n}{n-k} \right)^{n-k}} \]

according to inequality (A) with $n' = n$ and $t = \lambda / n$

\[e^{- \frac{\lambda^2}{n-k}} < \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]

\[e^{- \frac{k^2}{n-k}} = e^{- \frac{k^2}{n-k} \left( 1 - \frac{\lambda}{n} \right)^{n-k}} \]

Thus \[p(k; \lambda) . e^{-} \geq \frac{\lambda^{n-k}}{k_1!} e^{- \left( \frac{n}{n-k} \right)^{n-k}} < \frac{\lambda^{n-k}}{k_1!} \left( 1 - \frac{k}{n} \right)^{n-k} \]

\[p(k; \lambda) . e^{-} \geq b(k; n, p) \]

\[p(k; \lambda) . e^{-} \geq b(k; n, p) \]

\[p(k; \lambda) . e^{-} \geq b(k; n, p) \]

\[p(k; \lambda) . e^{-} \geq b(k; n, p) \]
Table 2.1 is similar to table 1.1. It gives the maximum errors, $M_3(np)$ for the Poisson approximation, $B_3(k; \lambda)$ to the binomial distribution, $B(k;n,p)$. This approximation has the interesting property that its maximum errors seem to be practically independent of $n$. They seem to depend only on the probability $p$, approaching 0 as $p$ decreases toward 0.

Table 2.1

<table>
<thead>
<tr>
<th>values of $p$</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
<td>.001</td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td>.008</td>
<td></td>
<td></td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.01</td>
<td></td>
<td>.003</td>
<td></td>
<td></td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
</tr>
<tr>
<td>.02</td>
<td></td>
<td>.006</td>
<td>.005</td>
<td>.005</td>
<td>.005</td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td>.04</td>
<td>.011</td>
<td>.009</td>
<td></td>
<td>.010</td>
<td></td>
<td></td>
<td>.010</td>
</tr>
<tr>
<td>.05</td>
<td></td>
<td>.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.012</td>
</tr>
<tr>
<td>.08</td>
<td></td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.020</td>
</tr>
<tr>
<td>.1</td>
<td>.025</td>
<td>.029</td>
<td>.027</td>
<td>.026</td>
<td>.026</td>
<td></td>
<td>.026</td>
</tr>
<tr>
<td>.2</td>
<td>.063</td>
<td>.052</td>
<td>.055</td>
<td>.055</td>
<td>.054</td>
<td></td>
<td>.054</td>
</tr>
<tr>
<td>.4</td>
<td>.125</td>
<td>.127</td>
<td>.124</td>
<td>.124</td>
<td>.123</td>
<td></td>
<td>.123</td>
</tr>
<tr>
<td>.5</td>
<td>.177</td>
<td>.172</td>
<td>.166</td>
<td>.167</td>
<td>.167</td>
<td></td>
<td>.166</td>
</tr>
</tbody>
</table>

JASA Vol. 51 pg. 299
Often times in binomial probability calculations we do not need the aid of an approximating function. When \( n \) does not exceed a certain value we can use exact Binomial tables. The two most useful tables are, Tables of the Binomial Probability Distribution published by the National Bureau of Standards, and the Tables of the Cumulative Binomial Probability Distribution published by the Harvard University Computation Laboratory. The former gives both individual terms and cumulated terms for values of \( p \) ranging from \( .01 \) to \( .5 \) (by increments of \( .01 \)) and for \( n \) ranging from 2 to 49. The latter gives cumulative terms for similar values of \( p \) and for values of \( N \) up to 1000 (by varying increments).

To get an approximate value of a cumulative binomial probability distribution, we may use normal tables for "large" values of \( n \), or we may use Poisson tables when we have a "small" value of \( p \). However, it is often true that \( p \) is not small enough to give really accurate results when Poisson tables are employed, while \( n \) is too small for accurate use of normal tables.

It frequently happens that an upper bound for \( P(r \leq X \leq s) \) would serve our purpose. We will show how to find this from Poisson tables with greater accuracy than could be obtained by using these tables in the way described earlier.

We shall denote the general term of the binomial expansion by:

\[
B_i = \frac{n!}{i!(n-i)!} p^i q^{n-i}
\]
and the general term of the corresponding Poisson distribution with the same value of p by: \[ P_i = (pn) \cdot e^{-pn} / i! \]

We shall also consider a second Poisson distribution whose general term is given by: \[ P'_i = (p'n) \cdot e^{-p'n} / i! \] where \( p' \neq p \) will be determined later.

We shall use the following notations:

- \( U_i = B_{i+1}/B_i = (n-1)p/(i+1) \cdot (1-p) \)
- \( V_i = P_{i+1}/P_i = pn/(i+1) \)
- \( V'_i = P'_{i+1}/P'_i = p'n/(i+1) \)
- \( U_i - V_i = p(np - i) / (i+1) \cdot (1-p) \)

From the last equation we obtain the following:

\( U_i > V_i \) or \( U_i < V_i \) according as \( i < np \) or \( i > np \).

Thus the size of the general term of the binomial expansion fall off more steeply to the right of \( i = np \) than does that of the general Poisson term. Using this fact, we can obtain an upper bound to \( P(r < X < s) \) for any \( r \rightarrow np \). In fact,

\[
\frac{B_r}{B_r} = \frac{P_r}{P_r} \quad B_r = \frac{B_r P_r}{P_r} ;
\]

\[
\frac{B_{r+1}}{B_r} < \frac{P_{r+1}}{P_r} \quad B_{r+1} < \frac{B_r P_{r+1}}{P_r} ;
\]

\[
\frac{B_{r+2}}{B_{r+1}} < \frac{P_{r+2}}{P_{r+1}} \quad B_{r+2} < \frac{B_{r+1} P_{r+2}}{P_{r+1}} < \frac{B_r P_{r+2}}{P_r} ;
\]

\[
B_s < \frac{B_s P_s}{P_r}.
\]

Adding these, we obtain

\[
P(r < X < s) = \sum_{i=r}^{s} P_i \cdot (B_i / P_i) \sum_{i=r}^{\infty} \frac{P_i}{i} = \left( \frac{P_r}{P_r} \right) \cdot \left( \sum_{i=r}^{\infty} P_i - \sum_{i=r}^{\infty} P_i \right) \quad (1)
\]

The quantity in parentheses can be found by use of the

16 AMS. Vol. 19, pg. 592
cumulative Poisson table provided, of course, it is within the range of that table, while the $B_r/P_r$ can be computed directly.

In the work we have done so far, we have used a Poisson distribution which is less steep than the corresponding binomial distribution throughout the whole interval, $np < r < X < n$. It seems reasonable to investigate the possibility of improving upon (1) by using a Poisson distribution having a different value $p'$ in place of $p$, where $p'$ is chosen so that the new Poisson distribution is of the same steepness at $X=r$ as is the binomial distribution. We wish to have $U_r = V'_r$ and $U_i < V'_i$ for all $r < i < n$. The first of these conditions requires that:

$$(n-r)p / (r+1)(1-p) = p' n / (r+1).$$

Solving for $p'$, we obtain

$$p' = (n-r)p / (n)(1-p).$$

Now we can prove the following:

If we have $p'$, $U'_i$, $V'_i$ and $V'_i$ as defined above, then

$$U'_i < V'_i < V'_i$$

provided $r > np$ and $i > r$.

since

$$U'_i / V'_i = (n-i)p(1+i) / (1+i)(1-p)np'$$

and this reduces to

$$U'_i / V'_i = (n-1) / (n-r)$$

by replacing $p'$ by its value above.

Then $U'_i / V'_i < 1$ since $i > r$. Thus $U'_i < V'_i$.

Moreover, we have

$$V'_i / V'_i = p'n(1+i) / (1+i)np = p' / p = (n-r) / (n-np),$$

but $r > np$ and hence $V'_i < V'_i$. 


Thus $U_i < V_i' < V_i$ where $r > np$ and $i > r$.

We now can obtain an inequality somewhat better than (1).

The derivation of the new upper bound for $P(r < X < s)$ goes just as before except that each $P_i$ is replaced by $P_i'$.

We obtain the new inequality:

$$ P(r < X < s) < K'B_r/P_r' $$

where

$$ K' = \sum_{i=r}^\infty P_i' - \sum_{i=s+1}^\infty P_i' $$

We can get a lower bound as well as a somewhat improved upper bound for $P(r < X < s)$ by calculating $B_r$ and $B_{r+1}$ directly and then applying (1) or (2) to find an upper bound $M$ of $P(r+1 < X < s)$. This gives the inequality

$$ B_r + B_{r+1} < P(r < X < s) < B_r + M. $$

This could, of course, be still further improved by calculating directly still more of the $B$'s and using a similar procedure, but this can become tedious.

To illustrate the various approximations, we will give the results of a numerical example: For convenience in checking, a value of $n$ which is within the range of the tables is used.

Take $s = n = 40$; $r = 10$; $p = 1/10$; $p' = 1/12$

From exact tables:

$$ P(10 < X < 40) = P(X > 10) - P(X > 41) = .0050631 $$

Poisson tables in the usual way:

$$ P(10 < X < 40) = P(10, 4) - P(41, 4) = .008132 $$

which is not particularly good.

Using inequality (1) we obtain:

$$ P(10 < X < 40) = .6790(.008132) = .005522 $$
Using inequality (3) and inequality (1) to obtain M:

\[ 0.004682 < P(10 \leq X \leq 40) < 0.003594 + 0.001607 = 0.00520 \]

We can obtain a still better result by using inequality (2) to obtain M and then use inequality (3):

\[ P(10 \leq X \leq 40) < 0.005087 \]
3. GRAM-CHARLIER APPROXIMATIONS

1) NORMAL GRAM-CHARLIER APPROXIMATION

Another method for approximating the binomial distrib. is to choose an approximating function from some class of functions such as the Pearson type distribution or the Gram-Charlier function.

We have a random variable $X$ which is the sum, $X=X_1 + X_2 + \ldots + X_n$ of $n$ Bernoulli variables. Under the condition of Laplace's theorem, the distribution function $F(X)$ of the standardized variable $X - \frac{np}{\sqrt{npq}}$ is for large $n$ approximately equal to $\Phi(X)$, the normal distribution. We can write:

$$F(X) = \Phi(X) + R(X)$$

this implies that $R(X)$ is small for large values of $n$, so that $\Phi(X)$ may be regarded as a first approximation to $F(X)$.

It is then natural to ask if, by further analysis of the remainder terms we can find more accurate approximations; e.g. in the form of some expansion of $R(X)$ in series.

We are going to discuss an expansion in orthogonal polynomials, known as the Gram-Charlier series. From which we will discuss what is known as the normal Gram-Charlier approximation, and the Poisson Gram-Charlier approximation.

Let us first consider the Gram-Charlier Series. This is a rather general system of distribution functions which is based upon the normal distribution and its derivatives. This distribution has been found satisfactory for fitting or

\[\text{Cramer pg. 221}\]
"smoothing" certain empirical distributions.

The "generator" of this series is the normal distribution.

Let 
\[ \phi_0(X') = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X')^2} \]

and let 
\[ \phi_i = \frac{\partial^i}{\partial (X')^i} \phi_0(X') \quad \text{where} \quad X' = X - a \quad \text{and} \quad E X = a \]

\[ \sigma(X) = \sigma \]

Then the Gram-Charlier Series is:

\[ f(X) = b_0 \phi_0(X') + b_1 \phi_1(X') + b_2 \phi_2(X') + \ldots \]

\[ \phi_1(X') = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X')^2} \]

\[ \phi_2(X') = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X')^2} \]

\[ \phi_3(X') = 2 \phi_0(X') X' + \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X')^2} \right) \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X')^2} \right) (X'^2 - 1) \phi_0(X') \]

Let 
\[ D = \frac{d}{dX'} \]

then the so-called Tchebycheff-Hermite polynomials, \( H_n(X') \), is given by the equation:

\[ (-D) \phi_0(X') = H_n(X') \phi_0(X') \]

Evidently \( H_n(X') \) is of degree \( n \) in \( X' \) and the coefficient of \( X' \) is 1. By convention \( H_0 = 1 \). We have:

\[ \phi_0(X' - T) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X'-T)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X'^2 - 2X'T + T^2)}{2}} \]

\[ = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X'^2 - 2X'T + T^2)}{2}} \]

and also by a Taylor's expansion around \( X' \).
\[
\phi_o(x' - t) = \phi_o(x') - \frac{T d\phi_o}{dx'} + \frac{T^2}{2!} \frac{d^2\phi_o}{dx'^2} + \ldots
\]

\[
= \phi_o(x') - T \frac{d\phi_o}{dx'} + \frac{T^2}{2!} \frac{d^2\phi_o}{dx'^2} + \ldots
\]

\[
= \sum_{j=0}^\infty \frac{(-1)^j}{j!} T^j \frac{d^j}{dx'^j} \phi_o(x') = \sum_{j=0}^\infty \frac{(-1)^j}{j!} H_j(x') \phi_o(x')
\]

\[
\phi_o(x') e^{-\frac{(x'^2 - t^2)}{2}} = \phi_o(x') \sum_{j=0}^\infty \frac{T^j}{j!} H_j(x')
\]

consequently \( H_n(x') \) is the coefficient of \( \frac{T^n}{n!} \) in \( e^{-\frac{(x'^2 - t^2)}{2}} \)

\( H_n(z) \) is found to be:

\[
Z^n - \frac{n(n-1)}{2} Z^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2 \cdot 4} Z^{n-4}
\]

\[
f(x') = \sum_{n=0}^\infty \sum_{m} b_{m,n} \phi_o(x') \phi_n(x')
\]

Using the orthogonal property, namely, that

\[
\int_{-\infty}^{+\infty} H_n(x') \phi_n(x') \phi'_o(x') \, dx' = 0 \quad m \neq n
\]

\[
= n! \quad m = n
\]

we can write down an expression for \( b_{m,n} \)

we have \( f(x') = \sum_{n=0}^\infty \sum_{m} b_{m,n} \phi_n(x') \phi'_o(x') \)

Multiplying by \( H_n(x') \) and integrating from \(-\infty\) to \(+\infty\) we have,

in virtue of the orthogonal relationship:

\[
b_n = \frac{i}{n!} \int_{-\infty}^{+\infty} f(x') \cdot H_n(x') \, dx'
\]

Substituting in the explicit value of \( H_n(x') \):

\[
b_n = \frac{i}{n!} \int_{-\infty}^{+\infty} f(x') \cdot (x^n - \frac{n(n-1)}{2} x^{n-2} \ldots) \, dx'
\]
\[ b_n = \frac{1}{n!} \left[ \mu'_n - n (n-2) \mu'_{n-2} + \cdots \right] \]

where \( \mu'_k \) is the \( k \)-th moment about the origin,

\( f(X') \) is the density function of the standardized variable \( X' \),

which has mean zero and unit standard deviation, while its

\( r \) th moment is:

\[ \frac{\mu'_r}{\sigma^r} = \mu'_r \]

\[ b_0 = \int_{-\infty}^{+\infty} f(X') \, dX' = 1 \quad b_1 = \mu'_1 = 0 \]

\[ b_2 = \frac{1}{2} (\mu'_2 - \mu'_0) = \frac{1}{2} (1 - 1) = 0 \]

\[ b_3 = \frac{1}{6} (\mu'_3 - 3 \mu'_1) = \frac{\mu'_3}{6 \sigma^3} \]

\( f(X') \) can now be written:

\[ f(X') = \phi_0(X') - \frac{\mu'_3}{3!} \frac{\phi_3(X')}{\sigma^3} \left[ (X')^3 - 3 X' \right] + \ldots \]

where \( \mu'_3 \) is the 3rd moment of \( f(X') \).

Now, let \( f(X') \) be a binomial distribution with mean = \( np \),

and variance = \( npq \).

The 3rd moment \( \mu'_3 \) is \( npq (q - p) \)

Thus \( f(X') = \phi_0(X') - \frac{(q - p)}{6 \sqrt{npq}} \frac{\phi_3(X')}{3!} \left[ (X')^3 - 3 X' \right] \)

\[ = \phi_0(X') - \frac{(q - p)}{6 \sqrt{npq}} \frac{\phi_3(X')}{3!} \left[ (X')^3 - 3 X' \right] \]

where \( X' = X - np \)

\[ \Phi(X') = \int_{-\infty}^{X'} f(X') \, dX' = \int_{-\infty}^{X'} \phi(X') \, dX' - \frac{(q - p)}{6 \sqrt{npq}} \int_{-\infty}^{X'} \frac{\phi_3(X')}{3!} \, dX' \]

\[ = \Phi(X') - \frac{q - p}{6 \sqrt{npq}} \phi_0(X') \frac{\phi_3(X')}{3!} \left[ (X')^3 - 3 X' \right] \]

This is the normal Gram-Charlier approximation.
There is a question about whether the Gram-Charlier expansions really converge and represent $F(X)$. In practical applications it is in most cases only of little value to know the convergence properties of our expansion. What we really want to know is whether a small number of terms suffices to give a good approximation to $F(X)$. If we know this to be the case, it does not concern us much whether the infinite series is convergent or divergent. It is possible that even when the infinite series diverges its first few terms will give an approximation of an asymptotic character. What we would really like to know is if the approximation would improve if we included more terms in the series. Actually we are dealing with a question relating to the asymptotic properties of our expansion for large value of $n$.

Let's call $B_1(k;n,p) = \Phi(Z) - (q-p)\Phi'(Z) - 1 = Q(X)$
where $\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-t^2} dt$ and $Z = \left( k+\frac{1}{2} - np \right) / \sqrt{npq}$

$B_1$ is the Normal Gram-Charlier approximation.

Table 3.1 is the counterpart of table 1.1 for the approx. $B_1$. The pattern for $M_4(np)$ are similar to those of $M_1(np)$ with the errors, however, much smaller. It can be shown that $M_4(np) \leq .056/\sqrt{npq}$ for all values of $n$ and $p$. 

Cramer pg. 223
### Table 3.1

Values of $p$ & 5 & 10 & 25 & 50 & 100 & 250 & 500 & $\infty$
---
.002 & .002 & & & & & .055 & 
.004 & .004 & .055 & .024 & & & & 
.008 & .008 & .055 & .024 & & & & 
.01 & .01 & .054 & .009 & & & & 
.02 & .02 & .054 & .023 & .009 & .004 & & & 
.04 & .04 & .052 & .022 & .004 & & & & 
.05 & .05 & .052 & .022 & & & & & 
.08 & .08 & .054 & .044 & .013 & .007 & .003 & & 
.11 & .11 & .054 & .044 & .013 & .007 & .003 & & 
.2 & .2 & .040 & .015 & .006 & .003 & .002 & & 
.3 & .3 & .020 & .009 & .003 & .002 & .001 & & 
.4 & .4 & .012 & .006 & .002 & .001 & .001 & & 
.5 & .5 & .011 & .005 & .002 & .001 & .001 & & 

Values of $np$

<p>| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 |
| .5 | .5 | .5 | .5 | .5 | .5 | .5 | .5 |
| .054 | .054 | .054 | .054 | .054 | .054 | .054 | .054 |
| 1.0 | .040 | .044 | .052 | .054 | .054 | .055 | .055 | .055 |
| 1.5 | .020 | .020 | .020 | .020 | .020 | .020 | .020 | .031 |
| 2.0 | .012 | .015 | .020 | .022 | .023 | .024 | .024 | .024 |
| 2.5 | .011 | .013 | .015 | .015 | .015 | .015 | .015 | .016 |
| 3.0 | .009 | .009 | .009 | .009 | .009 | .009 | .009 | .009 |</p>
<table>
<thead>
<tr>
<th>Values of $np$</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5</td>
<td>.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>.002</td>
<td>.003</td>
<td>.003</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.0</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>.001</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>25.</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ii) POISSON GRAM-CHARLIER APPROXIMATION

Similar to the normal case, we have what is called a Poisson Gram-Charlier approximation. Just as the normal Gram-Charlier series (sometimes known as type A) was derived from the normal distribution, a Poisson Gram-Charlier Series (sometimes known as a type B series) is derived from the Poisson distribution; \[ p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!} \]

The normal Gram-Charlier series is made up of the derivatives of the normal density function \( \phi \), whereas the Poisson Gram-Charlier series is made up of the differences of \( p(k; \lambda) \).

\[
f(K) = b' \nabla p + b'' \nabla^2 p + b''' \nabla^3 p + \ldots + \sum_{i=0}^{\infty} b_i \nabla^i p
\]

where \( \nabla p(k-1; \lambda) = p(k; \lambda) - p(k-1; \lambda) \).

\[
\nabla p(k-2; \lambda) = \nabla p(k-1; \lambda) - \nabla p(k-2; \lambda)
\]

\[
= p(k; \lambda) - 2p(k-1; \lambda) + p(k-2; \lambda)
\]

\[
= \frac{e^{-\lambda} \lambda^k}{k!} - \frac{2e^{-\lambda} \lambda^k}{(k-1)!} = -\frac{e^{-\lambda} \lambda^k}{k!} \left(1 - \frac{k}{\lambda}\right)
\]

\[
= p(k; \lambda)(1 - \frac{K}{\lambda})
\]

\[
\nabla p(k-2; \lambda) = \nabla p(k-1; \lambda) - \nabla p(k-2; \lambda)
\]

\[
= p(k; \lambda) - 2p(k-1; \lambda) + p(k-2; \lambda)
\]

\[
= \frac{e^{-\lambda} \lambda^k}{k!} \left[1 - \frac{2K}{\lambda} + \frac{K(K-1)}{\lambda^2}\right] = p(k; \lambda) \left(1 - \frac{2K}{\lambda} + \frac{K(K-1)}{\lambda^2}\right)
\]

\[
f(k) = p(k; \lambda) \left[ b' + b'1 \left(1 - \frac{K}{\lambda}\right) + b'2 \left(1 - \frac{2K}{\lambda} + \frac{K(K-1)}{\lambda^2}\right) + \ldots \right]
\]

\[
define\ polynomials\ G^\eta_n\ by\ the\ relation:\
\]
\[
p(k; \lambda) G^\eta_n(k; \lambda) = (-1)^n \nabla^n p(k-n; \lambda)
\]

Kendall, pg. 155
where

\[ G_0 = 1 \]

\[ G_1 = -(1 - \frac{\lambda}{\lambda}) \]

\[ G_2 = (1 - \frac{2\lambda}{\lambda} + \frac{\lambda(k-1)}{\lambda^2}) \]

etc.

\[ G_n(k;\lambda) = \frac{(-1)^n}{n!} \frac{n!}{p(k-n;\lambda)} \]

\[ G_n \]

can be calculated from the \( n \)th differences of the Poisson function, \( p(k;\lambda) \), in the same way that \( H_n \) may be derived from the \( n \)th differential coefficient of the normal distribution.

The \( G \)'s also obey the orthogonal law:

\[
\sum_{k=0}^{\infty} G_r^* G_s p(k;\lambda) = \begin{cases} r! & r = s \\ 0 & r \neq s \end{cases}
\]

we have

\[
f(k) = \sum_{j=0}^{\infty} b_j' \cdot (-1)^j G_j p(k;\lambda)
\]

\[
b_0' = \sum_{k=0}^{\infty} f_k G_0 = \sum_{k=0}^{\infty} f_k \cdot 1 = 1
\]

\[
b_0' = 1
\]

\[
b_1' = -\lambda \sum_{k=0}^{\infty} f_k = \lambda \sum_{k=0}^{\infty} (1 - \frac{\lambda}{\lambda}) = \lambda \sum_{k=0}^{\infty} f_k - \lambda \sum_{k=0}^{\infty} k \cdot f = \lambda - \lambda = 0
\]

since \( \sum_{k=0}^{\infty} k \cdot f = \lambda \)

\[
b_2' = 0
\]

\[
b_2' = \frac{\lambda^2}{2} \sum_{k=0}^{\infty} [1 - \frac{2\lambda}{\lambda} + \frac{\lambda(k-1)}{\lambda^2}] = \frac{\lambda^2}{2} \sum_{k=0}^{\infty} f_k - \lambda \sum_{k=0}^{\infty} f_k + \frac{\lambda}{2} \sum_{k=0}^{\infty} f_k^2 - \frac{1}{2} \sum_{k=0}^{\infty} k \cdot f
\]
Now, \( f(k) = p(k; \lambda) + \frac{3}{2}(\lambda - \lambda). \nabla p(k-2; \lambda) \)

\[ F(k) = p(k; \lambda) + \frac{3}{2}(\lambda - \lambda). \nabla p(k-1; \lambda) \]

\[ = p(k; \lambda) + \frac{3}{2}(\lambda - \lambda) \left( 1 - \frac{k}{\lambda} \right) p(k; \lambda) \]

For a binomial distribution \( \lambda = np \)

\[ F(k) = p(k; \lambda) + \frac{3}{2}(np - np). \left( \frac{np - k}{n} \right) p(k; \lambda) \]

\[ = p(k; \lambda) + \frac{3}{2} (k - np). p(k; \lambda) = B_{5}(k; \lambda) \]

\( B_{5}(k; \lambda) \) is the Poisson Gram-Charlier approximation.

Table 3.2 gives the maximum errors \( M_{5}(np) \) occurring when using the Poisson Gram-Charlier approximation to the binomial distribution. This approximation as in the Poisson approx. seems to have its maximum errors practically independent of \( n \). It depends only on the probability \( p \), approaching zero as \( p \) decreases toward zero. This approximation seems to be quite good even for fairly substantial values of \( p \).
Table 3.2

<table>
<thead>
<tr>
<th>Values of p</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.001</td>
<td>.002</td>
</tr>
<tr>
<td>.2</td>
<td>.007</td>
<td>.007</td>
<td>.006</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td>.3</td>
<td>.014</td>
<td>.016</td>
<td>.015</td>
<td>.016</td>
<td>.016</td>
</tr>
<tr>
<td>.4</td>
<td>.035</td>
<td>.030</td>
<td>.031</td>
<td>.031</td>
<td>.030</td>
</tr>
<tr>
<td>.5</td>
<td>.051</td>
<td>.052</td>
<td>.054</td>
<td>.053</td>
<td>.053</td>
</tr>
</tbody>
</table>
The Camp-Paulson approximation is based on an altogether different principle from those considered thus far. This approximation uses the approximation given by Paulson to the $\int$ of Snedecor's $F$ and the known fact that the $\int$ of $F$ is an Incomplete Beta function, and that the sum of terms of the binomial is also an Incomplete Beta function of suitable arguments.

The statistic $F = s_1^2 / s_2^2$, where $s_1^2$ and $s_2^2$ are two independent estimates of the same variance, has played an essential part in modern statistical theory. Paulson found a modified statistic $U$, a function of $F$, so selected as to tend to have a nearly normal distribution with 0 mean, and unit variance.

$F$ can be written in the form

$$F = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2}$$

where $\chi_1^2$ and $\chi_2^2$ are independent and have the chi-square distribution with $n_1$ and $n_2$ degrees of freedom respectively.

It is known from the work of Wilson and Hilferty that

$$\left( \frac{\chi_1^2}{n_1} \right)^{1/3}$$

is nearly normally distributed with mean, $1 - 2/9n$, and variance, $2/9n$.

Thus, Paulson approached the problem of securing an approximat.
to the \( F \) distribution by regarding \( \frac{Y}{X} \) as the ratio of two normally distributed variates. In general the distribution of the ratio \( V = Y / X \), where \( Y \) and \( X \) are normally and independently distributed with means, \( m_Y \) and \( m_X \), and standard deviations \( \sigma_Y \) and \( \sigma_X \), is not expressible in simple form. However, Fie\( \ddot{\text{u}} \)lhas has shown that a function \( R \) of \( V \), namely \( R = \frac{V \cdot m_X - m_Y}{\sqrt{\frac{\sigma_Y^2}{n_Y} + \frac{\sigma_X^2}{n_X}}} \), will be nearly normally distributed with zero mean and unit variance, provided the probability of \( X \) being negative is small. So that in the given problem it follows that we can regard

\[
U = \frac{1 - 2/9n_1}{F - (1 - 2/9n_1)}
\]

where \( V = F^{1/3} \), \( Y = \chi^2/n_1 \), \( X = \chi^2/n_2 \),

\[
m_Y = (1 - 2/9n_1) \quad m_X = (1 - 2/9n_2) \quad \sigma_Y^2 = 2/9n_1 \quad \sigma_X^2 = 2/9n_2
\]

as nearly normally distributed (0 mean, unit variance) provided \( n_2 > 3 \), this is to insure that the denominator is not negative. If we use the lower tail of the \( F \) distrib., then the statistic \( U \) should only be used if \( n_1 \) is also \( > 3 \).

We have yet to show that the summation of terms of a binomial is an incomplete Beta function, and that the \( \int \) of Snedecor's \( F \) is also an incomplete Beta function.

---

\(^{23}\) AMS Vol 24 pg. 428

\(^{24}\) AMS Vol 13 pg. 233
For Taylor's series with the integral form of remainder,

\[ f(a+h) = \sum_{j=0}^{r-1} \frac{h^j f^{(j)}(a)}{j!} + \int_0^h \frac{h^r (l-t)^{r-1}}{(r-1)!} f'(a+th) \, dt. \]

Putting \( a=q, \ h=p, \) and \( f(a+h) = (q+p) \) we have

\[ (q+p)^n = \sum_{j=0}^{r-1} \frac{n!}{j!(n-j)!} p^j + R_r \]

where \( R_r \) is the remainder after \( r \) terms and equals

\[ R_r = \int_0^p \frac{p^r (l-t)^{r-1}}{(r-1)!} \frac{n!}{(n-r)!} (q+pt)^{n-r} \, dt \]

In the last equation putting \( t = 1 - \frac{x}{p} \), we find

\[ R_r = \frac{n!}{(r-1)!} \int_0^1 x^{n-r} (1-x)^{r-1} \, dx \]

Beta function \( B(n_1, n_2) = \int_0^1 v^{n_1-1} (1-v)^{n_2-1} \, dv = \frac{\Gamma(n_1) \Gamma(n_2)}{\Gamma(n_1 + n_2)} \)

where \( \Gamma(n+1) = n! \) when \( n \) is an integer.

\[ R_r = \frac{\Gamma(n+1)}{\Gamma(n-r+1)} \cdot \frac{B(r,n-r+1)}{\Gamma(n-r+1)} \]

\[ = \frac{B(r,n-r+1)}{B(r,n-r+1)} \cdot \frac{\Gamma(n-r+1)}{\Gamma(n-r+1)} \]

\[ = \frac{\Gamma(n-r+1)}{\Gamma(n-r+1)} \cdot \frac{\Gamma(n-r+1)}{\Gamma(n-r+1)} \]

The remainder after \( k+1 \) terms (including the \( k+1 \) term) \( k \geq 0 \)

\[ \frac{\Gamma(k+1, n-k)}{\Gamma(n-k)} = \frac{B(k+1, n-k)}{B(k+1, n-k)} = F(x > k) \]

To show that the \( F \) of Snedecor's \( F \) is also an incomplete Beta function; we write \( F \) in the following form:

\[ \text{Wilks, pg. 75} \]

\[ \text{Kendall pg. 120} \]
The distribution of $F$ with $n_1$ and $n_2$ degrees of freedom can be shown to be

$$\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{n_1/2} \left(1 + \frac{n_1}{n_2} F\right)^{-\frac{n_1+n_2}{2}}$$

let's denote this by $h_{n_1,n_2}(F)$.

By a simple change of variable the $F$-distribution may be changed into a distribution which is the integrand of the Beta function times a constant.

let $X = \frac{n_1}{n_2} F \frac{1}{1 + \frac{n_1}{n_2} F}$ then $F = \frac{n_2 X}{n_1 (1-X)}$ and $dF = \frac{n_2}{n_1} \frac{dX}{(1-X)^2}$

so $h_{n_1,n_2}(F) \cdot dF$ transforms into

$$\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{n_1/2} X^{\frac{n_2}{2}-1} \left(\frac{X}{1-X}\right)^{\frac{n_2}{2}-\frac{n_1}{2}} \left(\frac{1}{1-X}\right)^{-\frac{n_1+n_2}{2}} dX$$

since

$$\left(1 + \frac{F}{n_2}\right)^{-\frac{n_1}{2}} = \left(1 + \frac{X}{1-X}\right)^{-\frac{n_1}{2}} = \left(\frac{1}{1-X}\right)^{-\frac{n_1-n_2}{2}} = \left(\frac{X}{1-X}\right)^{-\frac{n_1-n_2}{2}} \cdot X^{-\frac{n_1+n_2}{2}}$$

Wilks pg. 115
\[
\begin{align*}
\text{and } \quad (F) &= \left( \frac{n_2}{n_1} \frac{X}{1-X} \right)^{\frac{n_2}{2} - 1} \\
\int h_{n_1, n_2}(x) \, dx &= \int_0^X \frac{1}{B\left(\frac{n_1}{2}, \frac{n_2}{2} \right)} x^\left(\frac{n_1}{2} - 1\right) (1-x)^\left(\frac{n_2}{2} - 1\right) \\
&= \frac{B_x\left(\frac{n_1}{2}, \frac{n_2}{2} \right)}{B\left(\frac{n_1}{2}, \frac{n_2}{2} \right)} \\
&= I\left(\frac{n_1}{2}, \frac{n_2}{2} \right)
\end{align*}
\]

Comparing this with our result using the binomial:

\[
I_{\binom{k+1}{n-k}} x = p \quad \frac{n_1}{2} = k + 1 \quad \frac{n_2}{2} = n - k \quad n_1 = 2(k+1) \quad n_2 = 2(n-k)
\]

Thus \( F = \frac{n_2 X}{n_1 (1-X)} \)

can now be written: \( F = \frac{(n-k)p}{(k+1)q} \quad 1-p = q \)

We have from before

\[
U = \frac{1}{q} \left( q - \frac{2}{n_2} \right) F^{1/3} - \left( q - \frac{2}{n_1} \right)^{1/3} \sqrt{\frac{2}{n_2} F^{2/3} + \frac{2}{n_1}}
\]

which provides a normal approximation to the \( F \) distrib.

and if we now substitute the value of \( n_1, n_2, \) and \( F \)

that we have from above we get:

\[28\] Paulson, Vol. 13 pg. 235, AMS

Camp, AMS, Vol. 22 pg. 131
\[
U = \frac{1}{3} \left[ \left( \frac{1}{1-n} \right) \left[ \frac{(n-k)p}{(k+1)q} \right]^{1/3} - \left( \frac{1}{k+1} \right) \right]^{1/2}
\]

U is approximately normally distributed with mean 0, and variance 1.

Paulson's approximation seemed to be quite close. Since it was essentially an approximation to the incomplete Beta function, we must now have a similarly close approximation to the point binomial.

Thus, the Camp-Paulson approximation proceeded from the equivalence of a cumulative binomial probability to an incomplete Beta function, and thence to a probability integral of the variance ratio \( F \). Then we used an approx. to the integral of \( F \) developed by Paulson (who in turn used Wilson and Hilferty's approximation for the distribution of the chi-square and the result obtained by Fieller concerning the ratio of two normally distributed variates.) Then Camp developed the explicit expression which may be written

\[ B(k;n,p) \text{ is approximately equal to } \Phi(U) \]

where \( U \) is defined above.

Table 4.1 lists the maximum errors of the Camp-Paulson approximation. The error \( M_6(np) \) is strictly limited, with

AMS Vol. 22 pg. 131
an absolute maximum of .0122 which is never exceeded for any values of \( n \) and \( p \). It seems to be essentially a function of the mean \( np \), and tends to decline with increasing \( np \) when \( np > .02 \).

### Table 4.1

<table>
<thead>
<tr>
<th>Values of ( np )</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.02</td>
<td>.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.50</td>
<td>.005</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.002</td>
<td>.003</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
</tr>
<tr>
<td>1.5</td>
<td>.002</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.001</td>
<td>.003</td>
<td>.004</td>
<td>.005</td>
<td>.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>.001</td>
<td>.003</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>.001</td>
<td>.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>.001</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>.001</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

JASA, Vol. 51 pg. 300
BIBLIOGRAPHY


Hald, A. Statistical Theory with Engineering Applications (New York, John Wiley & Sons, Inc., 1952)


Camp, "Note on an Approximation to the Point Binomial" Annals of Mathematical Statistics (AMS) vol. 22, pg. 130

Cramer, "Note on an Approximation to The Binomial Summation" AMS vol. 19, pg. 592

Fieller, "The Distribution of the Index in a Normal Bivariate Population" AMS vol. 24, pg. 428

Paulson, "Note on an Approximation Normalization of the Analysis of the Variance Distribution" AMS vol. 13, pg. 233


Tukey and Freeman, "The transformation related to the Angular and the Square Root" AMS vol. 21, pg. 607
ABSTRACT

This thesis provides a discussion of various approximations to the cumulative binomial distribution.

We begin our analysis with the discussion of the simple normal approximation, based on the DeMoivre-Laplace Limit Theorem. This theorem states that the binomial distribution converges to the normal distribution in the situation wherein we hold $p$ constant and allow $n \to \infty$. The simple normal approximation is the most widely used of all approximations, because of its simplicity, and the availability of the necessary tables. However, its importance goes far beyond the domain of numerical calculation. We also show how the simple normal approximation may be used to obtain confidence intervals for the parameter $p$. We note various transformations that can be used to transform the binomial distribution into the normal distribution. In particular we mention the arcsine approximation which is based on the variance stabilizing angular transformation.

It is shown that in the binomial situation when the number, $n$, of trials approaches $\infty$, and the probability $p$, of success at each trial, approaches 0 in such a way that the variable $\lambda = np$ remains bounded, the Poisson distribution is an approximation to the binomial distribution. It is also noted that, when $\lambda = np$ is large, the Poisson distribution itself, can be approximated by means of the normal distribution. We also show that under certain conditions, we can get a
better approximation to the binomial distribution by using a modified Poisson approximation. This modified Poisson approximation uses a Poisson distribution having a different value for \( p \) then the binomial distribution.

Closely related to the simple normal approximation is the normal Gram-Charlier approximation. Here we derive the approximation starting from a Gram-Charlier type A series; that is, a series which is made up of the derivatives of the normal density function. This approximation is a little better than the normal approximation, because it adjusts for the skewness of the point binomial. Similarly, there is a Poisson Gram-Charlier approximation. This approximation is derived from a Gram-Charlier type B series, which is made up of the differences of the Poisson probability function. Results show, that this is one of the best approximations.

We conclude our discussion with the Camp-Paulson approx. This approximation is based on an entirely different principle from the previous ones. It turns out that the cumulative binomial probability is equivalent to an incomplete Beta function, and that the probability integral of the variance ratio \( F \) is also equivalent to an incomplete Beta function. By using an approximation to the integral of \( F \) derived by Paulson we obtain the Camp-Paulson approximation.

By comparing the various tables of the errors, \( M_{\xi}(np) \), we
can draw the following conclusions.

In terms of both accuracy and complexity, the approximations studied fall rather naturally into two groups. The less accurate "simple" approximations are the normal, the arcsine, and the Poisson. The "advanced" approximations are the normal Gram-Charlier, the Poisson Gram-Charlier, and the Camp-Paulson. The poorest of the "advanced" approximations is almost always more accurate than the best of the "simple" ones in any situation where it is appropriate.

Among the "simple" approximations the Poisson is generally the best when \( p \) is less than about 0.075. For larger values of \( p \) the arcsine approximation is usually the best, although the normal approximation overtakes it when \( p \) gets very close to one-half.

Among the "advanced" approximations, the Poisson Gram-Charlier is best in the same range that favors the Poisson among the "simple" ones. Everywhere else the Camp-Paulson approximation is best. The "poor" normal Gram-Charlier is never as good as the Camp-Paulson except in a small region where the Poisson Gram-Charlier is still better.

Hence, one needs only two approximations to match the cumulative binomial distribution almost exactly. For small values of \( p \) the Poisson Gram-Charlier approximation is exceedingly accurate; for larger values the same is true of the Camp-Paulson approximation. The maximum error can
be kept below .005, provided we exclude values of $n<5$ (where it is really unnecessary to use any approximations) by following the simple rule of using the Poisson Gram-Charlier approximation if $np < .8$ and the Camp-Paulson approximation if $np > .8$. 