The Incomplete Beta Function as a Contour Integral and a Quickly Converging Series for Its Inverse

M. E. Wise


Stable URL: http://links.jstor.org/sici?sici=0006-3444%28195012%2937%3A3%2F4%3C208%3ATIBFAA%3E2.0.CO%3B2-T

*Biometrika* is currently published by Biometrika Trust.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/bio.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.
THE INCOMPLETE BETA FUNCTION AS A CONTOUR INTEGRAL
AND A QUICKLY CONVERGING SERIES FOR ITS INVERSE

BY M. E. WISE

1. Introduction

Although so much work has been done on the incomplete beta function its mathematical treatment can still be simplified. In particular, there is a simple but quickly converging expansion for its percentage points which, surprisingly, seems to have been missed. It is for solving

\[ I\alpha(p, q) = \int_0^\alpha \frac{t^{p-1}(1-t)^{q-1} \, dt}{\int_0^1 t^{p-1}(1-t)^{q-1} \, dt} = P, \quad (1.1) \]

for either \( p \) or \( \alpha \), which are found in terms of percentage points of the \( \chi^2 \) distribution. The result was needed in a sampling problem for calculating \( p \) or \( \alpha \) in skew distributions in which \( p > q \), but it is found to be accurate even when \( p = q \), and \( \alpha \) is thereby obtained more easily than from the variance ratio 'F' or its logarithm \( z \).

The expansion is theoretically interesting in that it explains some other published approximations that depend on \( \chi^2 \). It is derived from a contour integral for \( I\alpha(p, q) \) which is expanded; with this approach it is easier to reverse the series, as it is not necessary to treat the numerator and denominator of (1.1) separately. We start instead from the incomplete negative binomial series and obtain also a simple proof of its well-known relation to the incomplete beta function and to the incomplete positive binomial series. This will be given first.

2. Incomplete binomial series and contour integrals

for the incomplete beta function

(a) The sum of the first \( p \) terms of the positive binomial series

In a random sample of \( n \) from a population with proportion \( x \) 'good' and \( 1-x \) 'bad', let \( Q \) be the probability that 0, 1, 2, \ldots or \( (p-1) \) are good, then

\[ Q = (1-x)^n + n(1-x)^{n-1}x + \frac{n(n-1)}{2!}(1-x)^{n-2}x^2 + \ldots + \frac{n(n-1)\ldots(n-p+2)}{(p-1)!}(1-x)^{n-p+1} \]

\[ = \frac{1}{2\pi i} \int_O (1-x+zx)^n \left( \frac{1}{z} + \frac{1}{z^2} + \ldots + \frac{1}{z^p} \right) \, dz, \quad (2.1) \]

where \( O \) is a contour of integration passing counter-clockwise round the origin

\[ = \frac{1}{2\pi i} \int_O (1-x+zx)^{n-p} \frac{dz}{1-z}, \quad (2.2) \]

always provided that \( z = 1 \) is outside of \( O \). To verify that this is the incomplete beta ratio, we find

\[ \frac{\partial Q}{\partial x} = -\frac{1}{2\pi i} \int_O n(1-x+zx)^{n-1}x^{p-1} \, dz. \]
Putting
\[
\frac{\partial Q}{\partial x} = \frac{n x^{n-1}(1-x)^{n-p}}{2\pi i} \int_{0}^{1} (1+z')^{n-1} z'^{n-p} dz'.
\]

(2.3)

\[Q = 0 \text{ when } x = 1 \text{ and } 1 \text{ when } x = 0, \text{ so}
\]
\[
Q = \int_{0}^{1} t^{n-1}(1-t)^{n-p} dt,
\]
and so
\[
1 - Q = I_2(n, p + q - 1).
\]

(2.4)

Karl Pearson (1924) first found this result by repeated integrations by parts of the beta integral.

(b) The sum of the first \( p \) terms of the negative binomial series

Later, Pearson gave the incomplete negative binomial series, also, as an incomplete beta function (1933), proved by a method found by E. C. Fieller from a real integral form of the remainder in Taylor's expansion (see also Kendall, 1945, p. 120). We shall now find the corresponding contour integral. If \(-q\) is the index of the series and \( Q' \) the sum of its first \( p \) terms,

\[
Q' = (1-x)^q \left\{ 1 + q x + \frac{q(q+1)}{2!} x^2 + \ldots + \frac{q(q+1)\ldots(q+p-2)}{(p-1)!} x^{p-1} \right\}
\]

(2.5)

\[
= \frac{(1-x)^q}{2\pi i} \int_{0}^{1} \left( \frac{1}{-ux^{p-1}} \right) \frac{du}{1-u}.
\]

(2.6)

Differentiating \( Q' \) with respect to \( 1/(1-x) \) and then putting \( u = x v \), it is easily found that

\[
\frac{\partial Q'}{\partial x} = \frac{x^{p-1}(1-x)^{p-1}}{2\pi i} \int_{0}^{1} u^{p-2}(1-u)^{1-q} du.
\]

(2.7)

Then since \( Q' = 0 \) when \( x = 1 \) and \( Q' = 1 \) when \( x = 0 \),

\[Q' = 1 - I_2(p, q).
\]

(2.8)

Alternatively, (2.6) can be written

\[
Q' = \frac{1}{2\pi i} \int_{0}^{1} \left\{ \frac{(1-x) u}{1-x u} \right\}^{p} \left( \frac{1-x}{1-x u} \right)^{q} \frac{du}{1-u}.
\]

(2.6′)

Now putting \( \frac{(1-x) u}{1-x u} = z \), (2.6′) reduces to (2.2), provided that

\[x = p + q - 1.
\]

(2.9)

(c) Generalization to any positive values of \( p \) and \( q \)

When \( p \) is not an integer (\( q \) of course can have any positive value) we define \( Q' \) by changing the contour of (2.6) into the one usually written as \( \int_{-\alpha}^{0+} \), which starts from \(-\infty - 0i\), goes
The incomplete beta function
round the origin in the positive sense and returns to \(-\infty + 0i\); in this case it must cross the real axis between 0 and 1. The corresponding path in the \(x'\) plane (equation (2-3)) is \(\int_{-1}^{0+i} \). It follows in the same way as before that
\[
Q' = Q'(x) = \frac{1}{2\pi i} \int_{-\infty}^{0+i} \frac{(1-xu)}{(1-u)}^{-1} u^{-p} du = 1 - I_\alpha(p, q),
\]
except that we have to prove that \(Q'(0) = 1\). This is easily done by changing the contour to one passing to the right of \(u = 1\); this adds \(-1\) to the integral. But the new path is equivalent to a circle with infinite radius, for which the integral \(Q'(0) - 1\) is clearly zero.

There is no need to evaluate the complete beta integral; this would come from (2-3), since
\[
\frac{1}{2\pi i} \int_0^\infty (1 + z')^{p-1} z'^{-p} dz' = \frac{(n-1)!}{(n-p)! (p-1)!},
\]
when \(p\) is a whole number; the generalization to all values of \(p\) is found in text-books (e.g. Whittaker & Watson, 1940, Chapter 12).

3. Outline of the Method of Obtaining the Expansions

The main idea was first applied by Molina (1932) to the beta integral
\[
\int_0^x t^{p-1} (1-t)^{q-1} dt.
\]

This is to express the integrand as a product of an exponential with a large index and a power series, and to integrate term by term, so that the resulting series is in negative powers of the index. Molina obtained a general expression for all possible series obtainable in this way, deduced as a special case (1932) a simple and quickly converging expansion of (3-1) in powers of \((p + \frac{1}{2} q - \frac{1}{2})^{-2}\), and used it to sum a few incomplete binomial series. The writer found the same series without finding the general expansion first (1946, 1948), in ignorance of Molina’s work, and deduced the corresponding expansion for the ratio \((1 : 1)\) (1946, equation (C 12)). In order to find this expansion directly without the laborious conversion of integral to ratio, we start from (2-10). The integrand is factorized in the same way, and the convergence is rapid because the factors can be chosen so that the power series is a hyperbolic sine (Wise, 1946, equation (G 8)). To do this, we use the identity
\[
1 - ux = (ux)^{1/2} \sinh \left( -\frac{1}{2} \log_\phi (ux) \right),
\]
and corresponding identities for \(1 - x\) and \(1 - u\), and put
\[ux = e^{-v}.
\]
Then
\[
1 - I_\alpha(p, q) = \frac{1}{2\pi i} \int_{-\infty}^{0} \frac{\sinh \frac{1}{2} y}{\sinh \left( -\frac{1}{2} \log_\phi x \right)} \left( e^{\frac{1}{2} pv + iv} \right)^{-2} e^{(p+1)v} e^{-\frac{1}{2} v \log_\phi x} dv,
\]
where \(c_\alpha\) goes from \(-\infty - \pi i\), across the real axis to the right of \(-\log_\phi x\), and back to \(-\infty + \pi i\).
This shows that the series is in powers of \(p + \frac{1}{2} q - \frac{1}{2}\). Write
\[
N = p + \frac{1}{2} q - \frac{1}{2},
\]
\[
y = -N \log_\phi x,
\]
and put
\[ w = N(v + \log_e x) = - N \log_e w, \]
then
\[ 1 - I_x(p, q) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{w} \left(1 + \frac{w}{2N}\right) \left(1 + \frac{w}{5760N^4}\right) \ldots}{\sinh \frac{w}{2N}} \frac{dw}{2N \sinh \frac{w}{2N}}, \tag{3.3} \]
which has an expansion in powers of \( N^{-1} \).

4. Expansion for the incomplete beta function in powers of \( (p + \frac{1}{2}q - \frac{1}{2})^{-2} \)

The expansion of (3.3) is
\[ 1 - I_x(p, q) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{w} \left(1 + \frac{w}{2N}\right) \left(1 + \frac{w}{5760N^4}\right) \ldots \frac{dw}{2N \sinh \frac{w}{2N}}, \tag{4.1} \]
where
\[ w_2(y) = w(q + 1) + 2qy, \]
\[ w_4(y) = 7w^3 + 9w^2 + (10q + 4)w + w^2(5q + 12). \]

It is convenient to write
\[ I_q = \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{w} \left(1 + \frac{w}{2N}\right)^{-q} \frac{dw}{2N \sinh \frac{w}{2N}}, \tag{4.2} \]
then putting \( w = t - y \)
\[ I_q = \frac{y^{q-1} e^{-y}}{2\pi i} \int_{-\infty}^{\infty} t^{-q} d\zeta = \frac{y^{q-1} e^{-y}}{(q - 1)!}, \]
since the integral in (4.2) is Hankel's contour integral for the reciprocal of the gamma function (see Whittaker & Watson, 1940, Chapter 12).

Write (4.1) as
\[ 1 - I_x(p, q) = 1 - P_0(y) - P_2(y) + P_4(y), \tag{4.3} \]
Now \( \frac{\partial P_0}{\partial y} = \frac{q}{y^2} I_{q+1} \), and \( P_0 = 0 \) when \( y \) is infinite and 1 when \( y \) is 0. These three facts prove that
\[ 1 - P_0(y) = \int_0^y y^{q-1} e^{-y} dy = \frac{\Gamma(q)}{\Gamma(q)}, \tag{4.4} \]

\( P_2 \) and \( P_4 \) are easily found by expressing \( w_2 \) and \( w_4 \) as polynomials in \( 1 + w/y \), and the ensuing integrals in terms of \( I_q \):
\[ P_2(y) = \frac{y^{q-1} e^{-y}}{(q - 2)!} (q + 1 + y), \tag{4.5} \]
\[ P_4(y) = \frac{y^{q-1} e^{-y}}{(q - 2)!} [(q - 3)(q - 2)(5q + 7)(q + 1 + y) - (5q - 7)y^2(q + 3 + y)]. \tag{4.6} \]

(4.3) is equivalent to the series obtained in the previous paper (Wise, 1946, equation (C12)). A similar result was obtained independently by Rao (1943).*

5. Expansions for the inverse function

If \( y \) is unknown and \( P \), which will be written for \( I_x(p, q) \), is known, clearly the first approximation to \( y \), \( y_0 \), say, is given by (4.4), i.e. (letting 1/\( N = 0 \))
\[ P = P_0(y_0), \tag{5.1} \]
or
\[ y_0 = \frac{1}{2} \chi_0(P), \tag{5.1'} \]

* I would like to thank Dr J. Wishart for drawing my attention to this work.
The incomplete beta function

i.e. half the value of \( \chi^2 \) exceeded by a proportion \( P \) of its distribution with \( 2q \) degrees of freedom. That is to say, if \( N \) is unknown and \( x \) is known, the first approximation \( N_0 \) to \( N \) is

\[
N_0 = \frac{x_0^2(P)}{2 \log_2 x}, \tag{5.2}
\]

and the first approximation to \( x \) for known \( N \) must be

\[
x_0 = \exp \left\{ \frac{-x_0^2(P)}{2N} \right\}. \tag{5.3}
\]

Thomson (1947) has found (5.3) by an interesting empirical argument, independently. His results will be discussed later.

The expansion for \( y \) as a function of \( P \) is clearly, if \( x \) is the unknown, of the form

\[
\frac{y}{y_0} = \log_2 x = 1 + \frac{\delta_2}{24N^2} - \frac{\delta_4}{5760N^4} + \ldots. \tag{5.4}
\]

Similarly, if \( N \) (i.e. \( p \)) is the unknown, we can put

\[
\frac{y}{y_0} = \frac{N}{N_0} = 1 + \frac{\delta'_2}{24N^2} - \frac{\delta'_4}{5760N^4} + \ldots. \tag{5.5}
\]

Then, since (5.4) and (5.5) are identical,

\[
\delta'_2 = \delta_2, \quad \delta'_4 = \delta_4 - 20\delta_2^2, \tag{5.6}
\]

and we only have to calculate \( \delta_2 \) and \( \delta_4 \). To do this, the right-hand side of (4.1) is expanded as a Taylor series about \( y = y_0 \). The resulting contour integrals are all expressed in terms of \( I_{q+3}, I_{q+2}, \) etc., by putting \( w = y_0(Y - 1) \). Then by using (4.2) they are expressed in terms of \( I_q \). Finally, equating coefficients of \( N^{-2} \) and \( N^{-4} \) gives

\[
\delta_2 = (q - 1)(q + 1 + y_0), \tag{5.7}
\]

\[
\frac{\delta_4}{(q - 1)(q + 1 + y_0)} = \frac{2y_0^2(4q - y_0 - 8)}{q + 1 + y_0} + (13q - 21)(q + 2) + 10(q - 1)(q + y_0 + \frac{1}{2}), \tag{5.8}
\]

and from (5.6)

\[
\frac{\delta'_4}{(q - 1)(q + 1 + y_0)} = \frac{2y_0^2(4q - y_0 - 8)}{q + 1 + y_0} + (13q - 21)(q + 2) - 10(q - 1)(q + y_0 + \frac{3}{2}). \tag{5.9}
\]

Thus the second approximation to \( x \) is given by

\[
\frac{y}{y_0} = \log_2 x = 1 + \frac{(q - 1)(q + 1 + y_0)}{24N^2}, \tag{5.10}
\]

and that for \( N \) by

\[
\frac{y}{y_0} = \frac{N}{N_0} = 1 + \frac{(q - 1)(q + 1 + y_0)}{24N^2_0}. \tag{5.10'}
\]

Adding the \( N^{-4} \) term multiplies the approximation to \( x \) by

\[
\exp \left\{ \frac{-y_0 \delta'_4}{5760N^4} \right\}, \tag{5.11}
\]

and adds to that for \( N \) the term

\[
\frac{\delta'_4}{5760N^4}. \tag{5.12}
\]
6. Numerical accuracy of the approximations given by (5.10), (5.10')

The error in \( x \) (or \( N \)) as calculated by neglecting \( \delta_q \) (or \( \delta_{pq} \)) is of interest. Since

\[
I_{x\sim}(q, p) = 1 - I_q(p, q),
\]

one can always make \( p \geq q \). In order to obtain the error as accurately as possible \( x \) has therefore been computed in the worst cases, i.e. \( p = q \), for comparison with some of the exact values published in Catherine Thompson's (1941) tables. The values of \( y_0 \) were obtained from Thompson's (1941) tables of \( x^2 \) percentage points; she mentions that linear interpolation is

Table 1. Percentage errors in \( x \) for \( p = q \)

<table>
<thead>
<tr>
<th>( q )</th>
<th>( N )</th>
<th>( P )</th>
<th>( x )</th>
<th>( \xi_p )</th>
<th>( \xi_q )</th>
<th>( \xi_{pq} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5-5</td>
<td>( y_0 )</td>
<td>0.996</td>
<td>2.5768</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exact</td>
<td>0.67221</td>
<td>1.74477</td>
<td>3.673265</td>
<td>6.6908</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage error</td>
<td>0.015</td>
<td>0.05</td>
<td>0.15</td>
<td>0.44</td>
</tr>
<tr>
<td>7.5</td>
<td>10-75</td>
<td>( y_0 )</td>
<td>1.53691</td>
<td>3.1519</td>
<td>5.6701</td>
<td>9.2747</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exact</td>
<td>0.86275</td>
<td>0.66355</td>
<td>0.5</td>
<td>0.33645</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage error</td>
<td>0.03</td>
<td>0.05</td>
<td>0.17</td>
<td>0.39</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>( y_0 )</td>
<td>6.9333</td>
<td>10.2996</td>
<td>14.6980</td>
<td>20.1280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exact</td>
<td>7.2933</td>
<td>0.6134</td>
<td>0.05</td>
<td>0.38365</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage error</td>
<td>0.05</td>
<td>0.095</td>
<td>0.18</td>
<td>0.33</td>
</tr>
<tr>
<td>30</td>
<td>44-5</td>
<td>( y_0 )</td>
<td>17.7573</td>
<td>23.2294</td>
<td>29.6673</td>
<td>37.1085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exact</td>
<td>6.6241</td>
<td>0.98260</td>
<td>0.5</td>
<td>0.4750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage error</td>
<td>0.07</td>
<td>0.12</td>
<td>0.18</td>
<td>0.28</td>
</tr>
<tr>
<td>60</td>
<td>89-5</td>
<td>( y_0 )</td>
<td>41.937</td>
<td>50.317</td>
<td>59.666</td>
<td>70.110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exact</td>
<td>41.620</td>
<td>0.55842</td>
<td>0.5</td>
<td>0.44158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage error</td>
<td>0.085</td>
<td>0.12</td>
<td>0.17</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Notes. The results are calculated from

\[
y = y_0 \left(1 + \frac{(q - 1)(q + 1 + y_0)}{2N^2} \right),
\]

where

\[
y = -N \log_2 x, \quad y_0 = \frac{1}{2}x(P), \quad N = p + \frac{1}{2}q - \frac{1}{2}.
\]

For a given \( q \) and \( P \) and varying \( p \), the percentage error is proportional to \( N^{-5} \). The calculated values of \( x \) are always larger than the exact values.

usually accurate enough for fractional values of \( q \). \( y_0 \) can also be found from Campbell's (1923) series, of which he has given eleven terms:

\[
y_0 = q + \frac{1}{2}q^2 + \frac{1}{2}(\xi^2_q - 1) + \frac{1}{2}(\xi^2_p - 7\xi_{pq})q^{-1} - \frac{1}{2}\left(16 - 7\xi^2_q - 3\xi^2_p\right)q^{-1},
\]

where

\[
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}t^2) dt = P.
\]

A recent paper by Riordan (1949) on this type of reversed series is also of interest.

Obviously for any fixed \( q \) and \( P \) the percentage error in \( x \), from (5.11), is nearly proportional to \((p + \frac{1}{2}q - \frac{1}{2})^{-5}\), i.e. to \( N^{-5}\); that for \( N \) from (5.12) to \( N^{-4}\). Therefore percentage errors have
been tabulated for \( p = q \), and also values of \( N^q \) and of \( N^q_0 \) for \( P = 0.5 \). Thus the errors for all other values of \( p(q) \) can be quickly estimated. The values of \( P \) selected for tabulation provide nearly equal intervals of \( \xi_p \).

Inclusion of the fourth degree terms appears to reduce the error in \( x \) or \( N \) by at least 90\% when \( p = q \); obviously it will reduce it still more when \( p > q \), since the sixth degree terms in (5.10) and (5.10') must change \( x \) or \( N \) by amounts proportional to \( zN^{-7} \) or \( NN_0^{-8} \) respectively, for fixed \( q \) and \( P \). Hence \( \delta_4, \delta_4' \) can be assumed to give this error. It can be expressed in powers of \( q^4 \) using (6.2). This gives

\[
\frac{\delta_4}{q-1} = 3(24q^4 + 21\xi_p q^4 + (11\xi_p^2 - 8)q^2 + \ldots), \tag{6.4}
\]

\[
\frac{\delta_4'}{q-1} = -8q^3 - 17\xi_p q^4 + \frac{1}{2}(8 - 59\xi_p^2)q^2 + \ldots. \tag{6.4'}
\]

Table 2. Percentage errors in \( N \)

<table>
<thead>
<tr>
<th>( q )</th>
<th>Exact ( N )</th>
<th>( P )</th>
<th>0.995</th>
<th>0.9</th>
<th>0.5</th>
<th>0.1</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5.5</td>
<td>( N_6 )</td>
<td>5.37</td>
<td>5.34</td>
<td>5.30</td>
<td>5.23</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>Percentage error</td>
<td>*</td>
<td>0.01</td>
<td>0.04</td>
<td>0.12</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>10.75</td>
<td>( N_6 )</td>
<td>10.47</td>
<td>10.42</td>
<td>10.34</td>
<td>10.54</td>
<td>10.10</td>
</tr>
<tr>
<td></td>
<td>Percentage error</td>
<td>-0.005</td>
<td>*</td>
<td>0.035</td>
<td>0.09</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>( N_6 )</td>
<td>21.38</td>
<td>21.28</td>
<td>21.15</td>
<td>21.01</td>
<td>20.83</td>
</tr>
<tr>
<td></td>
<td>Percentage error</td>
<td>*</td>
<td>0.015</td>
<td>0.03</td>
<td>0.06</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>44.5</td>
<td>( N_6 )</td>
<td>43.14</td>
<td>42.98</td>
<td>42.80</td>
<td>42.59</td>
<td>42.34</td>
</tr>
<tr>
<td></td>
<td>Percentage error</td>
<td>0.015</td>
<td>0.02</td>
<td>0.03</td>
<td>0.05</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>88.3</td>
<td>( N_6 )</td>
<td>88.65</td>
<td>88.38</td>
<td>88.10</td>
<td>85.80</td>
<td>85.50</td>
</tr>
<tr>
<td></td>
<td>Percentage error</td>
<td>0.015</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

* Error is between ±0.005%.

Notes. The results are calculated from

\[ N = N_6 + \frac{(q-1)(q+1+y_0)}{24N_6}. \]

The exact \( P \) is equal to \( q \) and for other values of \( p \) and a given \( q \) and \( P \) the percentage error is proportional to \( N_6^{-4} \). The calculated values are the larger except for \( q = 7,5, P = 0.995 \).

Thus for a value of \( P \) for which \( y_0 \approx q, \delta_4 \approx 72q^4 \), and then when \( p = q \) the factor multiplying into \( x \) is \( 1 - (2/1215) \). Obviously the percentage error in \( x \) increases quickly with \( \xi_p \) and therefore with \( 1 - P \). The absolute error increases also, and if \( P \) is small and \( p \) is only slightly greater than \( q \), one obtains a more accurate value of \( x \) by interchanging \( p \) and \( q \).

The exact values of \( x \) given in Table 1 are regarded as known in testing the accuracy of the calculation of \( N = N(x, q, P) \) (and hence of \( p \)). \( P \) is required in many binomial sampling problems, such as those of the authors quoted in §§8(a) and 8(b). In these problems it is often

† Prof. E. S. Pearson pointed this out and has given some illustrations; e.g. for \( P = 0.05 \) or less, \( p \) and \( q \) should be interchanged when \( p = 12, q = 10 \), but not when \( p = 15, q = 10 \).
useful to have also a mathematical formula for \( p \). Obviously the error in \( N \) is extremely small (e.g. if \( y_0 = q = p \), the \( N^1 \) term adds \(-N/3270\) to \( N \)). Thus the formula (5\( \cdot \)10') can be used even if \( p < q \); if \( p \ll q \) it would be better to interchange them and solve (4\( \cdot \)3) for \( q \) by iteration, but in most practical cases \( p > q \).

7. Comparison of errors in \( x \) with those from 'z' approximations

With \( p \geq 2q < 100 \) (say), the errors in \( x \) from (5\( \cdot \)10) are generally much smaller than those obtained from Cochran's (1932) formula.* However, Carter (1947) has recently found a more accurate formula of the same type, viz.

\[
x = \frac{1}{2q} \left( \frac{3}{s} - d \right),
\]

where \( \frac{1}{x} = 1 + \frac{q}{p} e^{2x}, \quad \lambda = \frac{1}{2} (\xi_2 - 3), \quad s = \frac{1}{2q - 1} + \frac{1}{2q - 1}, \quad d = \frac{1}{2q - 1} = \frac{1}{2q - 1}. \quad (7\cdot1)
\]

In deriving this approximation, a term \( \frac{1}{4q^2} \left( \frac{2}{s} \right) (\xi_2 + 11 \xi_2) \) is neglected.

For comparison, \( x \) has been calculated for \( p = 80, q = 15 \); in this case both approximations are good, but Carter's is much closer than (5\( \cdot \)10) at one end of the range of values of \( P \), whilst at the other end (5\( \cdot \)10) is slightly better:

<table>
<thead>
<tr>
<th>( P )</th>
<th>0.995</th>
<th>0.990</th>
<th>0.995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage error of ( x ) from (7\cdot1)</td>
<td>+0.010</td>
<td>+0.006</td>
<td>+0.003</td>
</tr>
<tr>
<td>Percentage error of ( x ) from (5( \cdot )10)</td>
<td>+0.004</td>
<td>+0.013</td>
<td>+0.012</td>
</tr>
</tbody>
</table>

A study of Carter's table of values of \( x \) calculated from (7\cdot1) and compared with the exact values (his \( n_t = 2q, n_s = 2p \)) shows that in all cases his percentage error increases slowly with \( P \) as well as with \( d \); that is to say, it changes with \( P \) and \( p/g \) in the opposite way to our errors—but more slowly. Thus doubling \( p \) has divided our error in \( x \) from (5\( \cdot \)10) for \( p = q \) by 14; trebling \( p \) (i.e. \( p = 3q \)) would divide it by 70. A rough working rule is that if \( p > 3q \), the \( \chi^2 \) formula is better than a 'z' approximation such as (7\cdot1); when \( 3q > p > 2q \) the 'z' formula is better for small \( P \) only, and for all \( P \) when \( p < 2q \). Clearly if the table of percentage points of \( \chi^2 \) could be extended to include more degrees of freedom, all incomplete beta percentage points could be calculated quickly and accurately; clearly, also, many of those already tabulated could have been computed much more easily.

8. Relation to other \( \chi^2 \)-type approximations

The expansion method provides the mathematical basis for several recently published approximations to the inverse function that depend on \( \chi^2 \).

(a) Scheffe & Tukey (1944) have given a simple approximation to \( p \) that is valid when \( x \) is between 0.9 and 1. It seems from a later comment (Birnbaum & Zukerman, 1949) that

* See also Thomson's table (1947, p. 371) under (D).
The incomplete beta function

it has not been proved mathematically. Their sample size \( n = p + q - 1 \) as in (2·1); thus in our notation \( N = n - \frac{1}{2}g + \frac{1}{2} \), and their formula is

\[
N = y_0 \left( 1 - \frac{1}{x} \right) \left( 1 - \frac{1}{y_0} \right). \tag{8·1}
\]

Thus \( N > 10y_0 \), so that the error in \( N \) from (5·10') is negligible; it is of the order of \( 1 \cdot 5 \times 10^{-2}N \). Expand (5·10') in powers of \( 1 - x = x' \) (say); this gives

\[
\frac{N}{y_0} - \frac{1}{x} + \frac{1}{2} = \left( 1 - \frac{1}{x} \right) \left( \frac{1}{y_0} \right) \left( 2 - (g-1)(g+1+y_0) \right) - \frac{x'^2}{2} \left( \frac{1}{x_0} \right) \left( 10(g-1)(g+1+y_0) \right), \tag{8·2}
\]

and the error in the Scheffé-Tukey approximation is \( y_0 \) times the right-hand side of (8·2). This explains why they find that the error is less than \( 1 \% \) when \( 1 - x < 0·1 \). To estimate this error more precisely we substitute Campbell's expansion (6·2) for \( y_0 \). This gives

\[
\frac{(g-1)(g+1+y_0)}{y_0^2} = 2 \left( 1 - \frac{3\xi}{2q^2} + \frac{5\xi^2}{3q^2} + \ldots \right), \tag{8·3}
\]

and

\[
\frac{N}{y_0} - \frac{1}{x} + \frac{1}{2} = \left( \frac{1}{x} \right) \left( \frac{1}{y_0} \right) \left( \xi \right) \left( 1 - \frac{5\xi}{2q^2} + \frac{5\xi^2}{3q^2} + \ldots \right). \tag{8·4}
\]

This is of interest also in explaining their second approximation to \( N \) which follows.

(b) MacCarthy (1947) has quoted a new approximation due to Scheffé & Tukey. In our notation it is

\[
N = g - \frac{(y_0 - g)x'}{1 - x'} - \frac{\xi}{2q}. \tag{8·5}
\]

This looks very different from (8·1) until it is expanded in powers of \( 1 - x \), when one gets

\[
\frac{N}{y_0} - \frac{1}{x} + \frac{1}{2} = \left( \frac{g}{y_0} - 1 \right) \left( \frac{1}{2} x' + \frac{5\xi^2}{3q} + \ldots \right). \tag{8·6}
\]

Substituting (6·2) for \( y_0 \),

\[
\frac{N}{y_0} - \frac{1}{x} + \frac{1}{2} = \left( \frac{g}{y_0} - 1 \right) \left( \frac{1}{2} x' + \frac{5\xi^2}{3q} + \ldots \right), \tag{8·7}
\]

\[
\left( \frac{N}{y_0} \right)_{\xi=0} - \left( \frac{N}{y_0} \right)_{\xi=0} = \left( \frac{2x' + x'^2}{144q} \right) - \frac{x'^3}{720} \left( \frac{1}{384q^2} + \frac{70\xi^2}{3456q} \right). \tag{8·8}
\]

Comparing this with the error (8·4) in their first approximation, the term in \( \xi g^{-4} \), which is the largest unless \( P \) is very near to \( \frac{1}{2} \), has been almost eliminated.

The remaining formulae are approximations to \( x \) and have been discussed by Thomson (1947).

(c) As already mentioned, the first approximation to \( x \), i.e. \( e^{-\xi y_0} \), corresponds to D. Halton Thomson's approximation (A); he has discussed and tabulated its error for \( P = 60 \) and \( 15 \) and various values of \( q \), for \( P = 0·995, 0·5 \) and \( 0·005 \). This is therefore a table of values of

\[
\frac{100(g-1)(g+1+y_0)y_0}{3(2p+q-1)^2}, \tag{8·9}
\]

when \( x < \frac{1}{2} \); for \( x > \frac{1}{2} \) his tabulated percentages are \( x/(1 - x) \) times this amount.
(d) Thomson’s approximation (B) to \( x \) is
\[
x = \left( \frac{p - \frac{1}{2}}{p + q - \frac{1}{2}} \right)^{\frac{g}{2q}}.
\]

This may be written as
\[
\log_2 x = \frac{2N}{q} \tanh^{-1} \left[ \frac{q}{2N + 1} \right] + \frac{q^2}{12N^2}.
\]

Comparing this with (5·10) the error of (8·11) is nearly equal to
\[
\frac{q^2}{12N^2} \left( \frac{q - 1}{24N^2} \right)
\]
and so is positive when \( P \) is near 1, negative when \( P \) is near zero, and small when \( P = \frac{1}{2} \) and \( y_0 = q \) as found by Thomson.

(e) Campbell (1923) found the expansion for \( x \) in powers of \( n = p + q - 1 \). He interpreted his first approximation to \( x \) as follows: Starting from the other end of the binomial series (2·1), the probability that there are \( q - 1 \) or less ‘bad’ in the sample of \( n \) is the sum of the first \( q \) terms, and, as is well known, if the average number of ‘bad’ in a sample of \( n \) is \( a \) and remains fixed whilst \( n \to \infty \), then \( x = 1 - a/n \) and
\[
P \to e^{-a} \left( 1 + a + \frac{a^2}{2!} + \ldots + \frac{a^{q-1}}{(q-1)!} \right) = 1 - \frac{\Gamma_c(q)}{\Gamma(q)}.
\]

So as \( n \to \infty, a \to y_0 \) and can be expanded in powers of \( n^{-1} \). In our notation, his
\[
A \sim \frac{1}{2} \left( q - 1 - y_0 \right) \quad (c = q),
\]
and his expansion is
\[
a = n(1 - x) = y_0 \left[ 1 + \frac{A}{n} + \frac{14A^2 + (3y_0 + 2)A + y_0}{12n^2} + \ldots \right.
\]
\[
\left. + \frac{36A^3 + (20y_0 + 12)A^2 + (3y_0^2 + 8y_0)A + y_0^2}{24n^3} + \ldots \right].
\]

This series is obtained from (5·10) by expanding \( x \) in powers of \( n^{-1} (N = n - \frac{3}{2}q + \frac{1}{2}) \). The term in \( n^{-3} \) has been obtained in Campbell’s series by Riordan. It is not obvious why one should instead get a better formula by expanding \( \log_2 x \) in powers of \( p + \frac{q}{2} - \frac{1}{2} \); had it been so, the new series might have been found 27 years ago.

**SUMMARY**

Expansions for percentage points of the incomplete beta function satisfying \( I_p(p, q) = P \), for either \( p \) or \( x \), have been found in terms of percentage points of the \( \chi^2 \) distribution. They are derived from a new contour integral for \( I_p(p, q) \) which is shown to be related to contour integrals for incomplete positive and negative binomial series. The latter integrals provide simple proofs of the known fact that these series are incomplete beta functions. The resulting approximations are good over all values of \( P \), \( q \) and \( p \) or \( x \), unless (for \( p \) unknown) \( p \leq q \), and they are extremely good for small \( 1 - P \) and/or small \( q/p \). They have been compared

* Unpublished. I would like to thank Mr. Riordan also for informing me of the work by SchefT and Tukey and MacCarthy.
The incomplete beta function

(i) numerically with exact tabulated values and with \( z \) approximations, and (ii) mathematically with some recently published approximations depending on \( \chi^2 \), which are thereby provided with a theoretical basis.

The author thanks Mr G. Klein (Birkbeck College, University of London) for a helpful mathematical discussion. The work was done while he was in the Physics Department of Birkbeck College, and he is indebted to his employers in the Philips Companies,\(^*\) and in particular to Dr J. A. M. van Moll, Head of the Material Research Laboratory, Mitcham, Surrey, for their support during this period.

REFERENCES


PEARSON, KARL (1934). Note on the relationship of the incomplete beta function to the sum of the first \( p \) terms of the binomial \((a + b)\)\(^*\). *Biometrika*, 16, 309.

PEARSON, KARL & FLIEDER, E. C. (1933). On the applications of the double Bessel function \( \kappa_{x,y}(x) \) to statistical problems. *Biometrika*, 25, 158.


* He is now employed by the Philips Research Laboratories, Eindhoven (Netherlands).