CONTRIBUTIONS TO STATISTICS

PRESENTED TO
PROFESSOR P. C. MAHALANOBIS
ON THE OCCASION OF HIS
70TH BIRTHDAY

ESTIMATION FOR THE GENERALIZED POWER SERIES
DISTRIBUTION WITH TWO PARAMETERS AND ITS
APPLICATION TO BINOMIAL DISTRIBUTION

By
GANAPATI P. PATIL

Paper received: March, 1963

PERGAMON PRESS
OXFORD • LONDON • NEW YORK • PARIS • FRANKFURT
STATISTICAL PUBLISHING SOCIETY
CALCUTTA
ESTIMATION FOR THE GENERALIZED POWER SERIES DISTRIBUTION WITH TWO PARAMETERS AND ITS APPLICATION TO BINOMIAL DISTRIBUTION*

By GANAPATI P. PATIL
McGill University

1. Introduction

Let \( T \) be a non-null countable subset with no limit point of real numbers and define the series function \( f(\theta) = \sum a_\theta \theta^x \), the summation extending over \( T \), with \( a_\theta > 0, \theta > 0 \) so that \( f(\theta) \) is positive, finite and differentiable. Then the distribution of a random variable \( X \) taking values in \( T \) with probabilities

\[
p(x; \theta) = \text{prob} \{X = x\} = \frac{a_\theta \theta^x}{f(\theta)} \quad x \in T
\]

is called a generalized power series distribution (GPSD). The author (1961, 1962a, 1962b, 1962c, 1962d) has studied some problems of statistical inference associated with the GPSD with parameter \( \theta \) given by (1). In the present paper, we consider a GPSD with two parameters taking the form

\[
p(x; \theta, \lambda) = \text{prob}\{X = x\} = \frac{a_\lambda(\lambda) \theta^x}{f(\theta, \lambda)} \quad x \in T
\]

where \( f(\theta, \lambda) = \sum a_\lambda(\lambda) \theta^x \), the summation extending over \( T \) such that \( f(\theta, \lambda) \) is positive, finite and differentiable for all admissible values of the two parameters \( \theta \) and \( \lambda \) and the positive coefficients \( a_\lambda(\lambda) \) now depend on \( x \) and \( \lambda \). The binomial and negative binomial distributions are special cases of the GPSD (2) when they are considered to be the distributions with two parameters.

2. Estimation of the two parameters for the GPSD

To estimate \( \theta \) and \( \lambda \) on the basis of a sample \( x_i \) \((i = 1, 2, \ldots, N)\) of size \( N \) from (2), the logarithm of likelihood function is

\[
\log L = \text{constant} + \sum_{i=1}^{N} x_i \log \theta + \sum_{i=1}^{N} \log a_{x_i}(\lambda) - N \log f(\theta, \lambda).
\]

The "efficient score" for \( \theta \) is then

\[
\Psi_1 = \Psi_1(\theta, \lambda) = \frac{\partial}{\partial \theta} (\log L) = \frac{N}{\theta} [\bar{x} - \mu(\theta, \lambda)]
\]

where \( \mu(\theta, \lambda) = E(X) \).

*The results of this paper form a part of the work carried out by the author while he was at the Indian Statistical Institute and the University of Michigan.
and the likelihood equation \( \Psi_x(\theta, \lambda) = 0 \) reduces to

\[
x = \mu(\theta, \lambda)
\]  \( \ldots \) (4)

which is the same as the first-moment equation.

The "efficient score" for \( \lambda \) is

\[
\Psi_{\lambda} = \Psi_{\lambda}(\theta, \lambda) = \frac{\partial}{\partial \lambda} \log L.
\]

\[
= N \left[ \sum_{i=1}^{n} \frac{d}{d\lambda} \log a_{x_i}(\lambda)/N \right] \frac{\partial}{\partial \lambda} \log f(\theta, \lambda)
\]  \( \ldots \) (5)

and the estimating equation \( \Psi_{\lambda}(\theta, \lambda) = 0 \) becomes

\[
\frac{d}{d\lambda} \log f(\theta, \lambda) = \frac{N}{\Sigma} \frac{d}{d\lambda} \log a_{x_i}(\lambda)/N.
\]  \( \ldots \) (6)

This, however, is not a moment equation. The second moment equation will be

\[
S^2 = \mu_{2}(\theta, \lambda)
\]

where

\[
S^2 = \Sigma \frac{(x_i - \bar{x})^2}{N}.
\]  \( \ldots \) (7)

Thus, unlike GPD's of the form (1) with single parameter, GPD's given by (2) with two parameters do not yield identical "moment" and "maximum likelihood" estimates.

The elements of the "information matrix"

\[
s = \left( \begin{array}{cc}
I_{11} & I_{12} \\
I_{12} & I_{22}
\end{array} \right)
\]  \( \ldots \) (8)

are given by

\[
I_{11} = -E \left( \frac{\partial \Psi_1}{\partial \theta} \frac{\partial \Psi_1}{\partial \theta} \right) = \frac{N}{\theta} \left( \frac{\partial \mu}{\partial \theta} \right)
\]

\( \ldots \) (9)

\[
I_{12} = -E \left( \frac{\partial \Psi_1}{\partial \lambda} \frac{\partial \Psi_2}{\partial \theta} \right) = N \left( \frac{\partial \mu}{\partial \lambda} \right)
\]

\( \ldots \) (10)

\[
I_{22} = -E \left( \frac{\partial \Psi_2}{\partial \lambda} \frac{\partial \Psi_2}{\partial \lambda} \right) = N \left[ \frac{\partial^2}{\partial \lambda^2} \log f(\theta, \lambda) - h(\lambda) \right]
\]

\( \ldots \) (11)

where

\[
\hbar(\lambda) = E \left[ \frac{d}{d\lambda} \log a_x(\lambda) \right] = \frac{1}{f(\theta, \lambda)} \left[ \frac{\partial^2}{\partial \lambda^2} f(\theta, \lambda) \right] - E \left[ \frac{d}{d\lambda} \log a_x(\lambda) \right]^2.
\]

The asymptotic "dispersion matrix" of the estimates \( \theta, \lambda \) obtained by solving (4) and (6) is then given by

\[
\begin{pmatrix}
\text{var} (\theta) & \text{cov} (\theta, \lambda) \\
\text{cov} (\theta, \lambda) & \text{var} (\lambda)
\end{pmatrix} = s^{-1}.
\]

\( \ldots \) (12)

If instead of \( \theta \) and \( \lambda \), \( \mu = \mu(\theta, \lambda) \) and \( \lambda \) are regarded as the parameters, the maximum likelihood estimates of \( \mu \) and \( \lambda \) are asymptotically uncorrelated. This follows from (10).
3. Estimation of the two parameters for the binomial distribution

The GPD (2) becomes

\[
p(x; \theta, \lambda) = \binom{\lambda}{x} \theta^x (1+\theta)^{\lambda-x} \quad x = 0, 1, 2, \ldots, \lambda
\]  \hspace{1cm} \text{... (13)}

when \( f(\theta, \lambda) = (1+\theta)^\lambda \). Writing \( \theta = \eta/(1-\eta) \) and \( \lambda = n \), (13) gives the binomial probability function in the well-known form

\[
p(x; \eta, n) = \binom{n}{x} \eta^x (1-\eta)^{n-x} \quad x = 0, 1, 2, \ldots, n
\]  \hspace{1cm} \text{... (14)}

for which one has \( \mu = \eta a \) and \( \rho = \eta (1-n) \).

The binomial distribution has essentially two parameters \( \eta \) and \( n \) of which \( n \) is usually known and only \( \eta \) has to be estimated. However, certain cases might arise in which \( n \) is unknown, and both \( n \) and \( \eta \) have to be estimated. For instance, while experimenting with a radioactive substance, in addition to the mean number \( (\mu = \eta a) \) of disintegrating atoms, it may perhaps be of interest to know the number \( (n) \) of atoms capable of disintegration for the substance in fixed intervals of time for some specified solid angle and fit a model correspondingly.

To estimate \( \eta \) and \( n \) on the basis of a random sample of size \( N \) with observed frequency \( n_a \) for \( x (\sum n_a = N) \) drawn from (14) with \( n \) unknown, the moment-estimates are given by

\[
x = \eta a
\]  \hspace{1cm} \text{... (15)}

and

\[
S^2 = \eta n(1-\eta)
\]  \hspace{1cm} \text{... (16)}

where \( x = \sum n_a/N \) and \( S^2 = \sum n_a (x-a^2)/N \).

The likelihood equations reduce to

\[
x = \eta \frac{\sum n_a}{N}
\]  \hspace{1cm} \text{... (17)}

and

\[
\sum \frac{T_{r+1}}{\eta^{r+1}} + N \log (1-\eta) = 0
\]  \hspace{1cm} \text{... (18)}

where

\[
T_{r+1} = \sum_{x=r}^n n_a.
\]  \hspace{1cm} \text{... (19)}

Eliminating \( \eta \) from (17) and (18) we have to solve for \( n \), the equation:

\[
\sum \frac{T_{r+1}}{\eta^{r+1}} + N \log \left[1-\frac{r}{\eta}\right] = 0.
\]  \hspace{1cm} \text{... (19)}

The elements of the information matrix are:

\[
I_{11} = nN/\pi(1-\pi)
\]

\[
I_{12} = N/(1-\pi)
\]

\[
I_{22} = E[\Sigma T_{r+1}/(n-r)^2] = N\Sigma [1-B(r; \pi, n)]/(n-r)^2
\]  \hspace{1cm} \text{... (20)}

where \( B(r; \pi, n) \) is defined by (23). We note that \( n \) is a discrete parameter and also the range of (14) depends on \( n \). The exact properties of the estimates are, therefore, not known.
4. Estimation for the Truncated Binomial and Incomplete Dibeta and Tribeta Functions

The applications of truncated binomial distributions have been discussed by several authors. The probability function of a truncated binomial with truncation points, say, at $c$ and $d$ can be written as:

$$b^*(x; \pi, n) = [B^*(c, d; \pi, n)]^{-1} \cdot \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

with

$$c = c, c+1, \ldots, d, \quad 0 \leq c < d \leq n$$

where

$$B^*(c, d; \pi, n) = B(d; \pi, n) - B(c-1, \pi, n)$$

and

$$B(r; \pi, n) = \sum_{x=0}^r b(x; \pi, n)$$

which can be written as the incomplete beta function

$$I_{1-n}(u-r, r+1) = \int_0^{1-n} u^{r-1}(1-u)^{d-1} du / \int_0^1 u^{r-1}(1-u)^{d-1} du.$$ 

The first two moments about the origin of (21) can be obtained as

$$\mu^* = \mu^*(c, d; \pi, n)$$

$$= n\pi \cdot B^*(c-1, d-1; \pi, n-1) / B^*(c, d; \pi, n)$$

and

$$m^*_2 = m^*_2(c, d; \pi, n)$$

$$= \mu^*(c, d; \pi, n) \cdot [1 + \mu^*(c-1, d-1; \pi, n-1)].$$

It is easy to see that a truncated GPD is a GPD in its own right and hence the truncated binomial is also a GPD and the results obtained for (2) apply to (21).

To estimate $\pi$ and $n$ on the basis of a random sample of size $N$ with observed frequency $n_\text{x}$ for $x$ ($\Sigma n_\text{x} = N$) drawn from the truncated binomial, (21), the moment equations are:

$$x = \mu^*$$

and

$$s^2 = m^*_2$$

where $x = \Sigma n_\text{x}/N$, $s^2 = \Sigma x^2 n_\text{x}/N$ and $\mu^*$ and $m^*_2$ are defined by (25) and (26) respectively.

The "efficient scores" for $\pi$ and $n$ reduce to:

$$\Psi_1 = \frac{N}{\pi (1-\pi)} (x - \mu^*)$$

and

$$\Psi_2 = \sum_{r \geq 0} \frac{N \cdot r + 1}{n - r} \cdot N \log (1-\pi) - N \frac{\partial B^*}{\partial n} \left/B^* \right.$$
The likelihood equations then become

\[ \bar{x} = \hat{\mu}^* \quad \ldots \quad (31) \]

and

\[ \Phi_2 = 0. \quad \ldots \quad (32) \]

The elements of "information matrix" are

\[ I_{11} = \frac{N}{\pi(1-\pi)} \cdot \frac{\partial \hat{\mu}^*}{\partial \pi} \]

\[ I_{12} = \frac{N}{\pi(1-\pi)} \cdot \frac{\partial \hat{\mu}^*}{\partial n} \]

and

\[ I_{22} = \left[ \frac{\partial^2}{\partial n \partial n^*} B^* | B^* - \left( \frac{\partial}{\partial n} B^* | B^* \right)^2 \right]_{B} \Sigma \frac{T_{r+1}}{(w-r)^2} \quad \ldots \quad (33) \]

(31) and (32) may be solved for estimation, approximating

\[ \frac{\partial}{\partial n} B(r; \pi, n) \sim \frac{\Delta B}{\Delta n} \]

and getting \( \frac{\Delta B}{\Delta n} \) from binomial tables where \( B \) is defined by (23).

However, exact values of \( \frac{\partial}{\partial n} B(r; \pi, n) \) and \( \frac{\partial^2}{\partial n^2} B(r; \pi, n) \) which we shall call "Incomplete Dibeta and Tribeta Function" respectively, can be obtained, as follows:

\[ \frac{\partial}{\partial n} B(r; \pi, n) = \frac{\partial}{\partial n} I_{1-n}(n-r, r+1) \]

\[ = I_{1-n}(n-r, r+1)[E_d(n-r, r+1)]_{n=1}^{w} \quad \ldots \quad (34) \]

where \( I \)'s are incomplete beta functions, and

\[ E_d(n, m) = \int_0^z \frac{\log u}{u^{m-1}(1-u)^{n-1}} \, du \]

\[ \ldots \quad (35) \]

which means the expected value of \( \log u \) when \( u \) follows a beta distribution truncated on the right at \( z \), with parameters \( n \) and \( m \). \( E_d(n, m) \) can be reduced to

\[ E_d(n, m) = \log z \frac{1}{I_z(n, m)} \sum_{r=0}^{m-1} \frac{I_x(n+r, m-r)}{n+r} \]

\[ \ldots \quad (36) \]
In particular,
\[ E_1(n, m) = - \sum_{r=0}^{m-1} \frac{1}{n+r}. \] ... (37)

(36) suggests that the values of "Incomplete Tribeta Function" can be exactly obtained by using tables of "Incomplete Beta Function" which are extensively tabulated.

To obtain "Incomplete Tribeta Function"

\[ \frac{\partial^2}{\partial n^2} B(r; \pi, n) = \frac{\partial}{\partial n} I_{1-r}(n-r, r+1)[E_4(n-r, r+1)]_{z=1}^{z=1-n} \] ... (38)

we get after some simplification of (38),

\[ \frac{\partial^2}{\partial n^2} B(r; \pi, n) = I_{1-r}(n-r, r+1)(([E_4(n-r, r+1)]_1^{1-r})^2 + [V_4(n-r, r+1)]_1^{1-r}) \] ... (39)

where \( V_4(n, m) \) is the variance of \( \log u \) when \( u \) follows a beta distribution with parameters \( n \) and \( m \) truncated on the right at \( z \). \( V_4(n, m) \) can be obtained from

\[ V_4(n, m) = E_4^2(n, m) - [E_4(n, m)]^2 \] ... (40).

where \( E_4^2(n, m) \) can be reduced to

\[ E_4^2(n, m) = (\log z)^2 - \frac{2}{I_4(n, m)} \sum_{r=0}^{n-1} I_4(n+r, m-r) \frac{E_4(n+r, n-r)}{n+r}. \] ... (41)

In particular,

\[ E_4^2(n, m) = 2 \sum_{r=0}^{m-1} \left( \sum_{r=0}^{n-1} \frac{1}{n+r+1} \right) \left( \frac{1}{n+r} \right). \] ... (42)

5. ILLUSTRATIVE EXAMPLES

The computation procedure for simultaneous estimation for the complete binomial (14) will be illustrated with reference to two examples: one on radioactive disintegrations and the second one on throwing of dice.

Example 1: The first two columns of the following table give data collected by Rutherford and others, showing the number \( n_x \) of intervals of time, each of 7.5 seconds, during which the number \( x \) of \( \alpha \) particles emitted from a certain radioactive substance.
GENERALIZED POWER SERIES DISTRIBUTION

TABLE I. DATA: RUTHERFORD AND GEIGER: RADIOACTIVE DISINTEGRATION

<table>
<thead>
<tr>
<th>number of α particles</th>
<th>number of intervals</th>
<th>$T_x = \Sigma r \cdot n_x$</th>
<th>$T_x \over n = x + 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>$n_x$</td>
<td>$T_x$</td>
<td>n=77</td>
</tr>
<tr>
<td>0</td>
<td>57</td>
<td>2608</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>203</td>
<td>2651</td>
<td>33.1299</td>
</tr>
<tr>
<td>2</td>
<td>383</td>
<td>2348</td>
<td>30.8947</td>
</tr>
<tr>
<td>3</td>
<td>525</td>
<td>1985</td>
<td>26.0000</td>
</tr>
<tr>
<td>4</td>
<td>532</td>
<td>1440</td>
<td>19.4504</td>
</tr>
<tr>
<td>5</td>
<td>408</td>
<td>908</td>
<td>12.4384</td>
</tr>
<tr>
<td>6</td>
<td>273</td>
<td>800</td>
<td>6.9444</td>
</tr>
<tr>
<td>7</td>
<td>139</td>
<td>227</td>
<td>3.1972</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>88</td>
<td>1.2671</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>43</td>
<td>0.6232</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>16</td>
<td>0.2363</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>6</td>
<td>0.0886</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>0.0303</td>
</tr>
<tr>
<td>total</td>
<td>2608</td>
<td>134.2995</td>
<td>131.0065</td>
</tr>
</tbody>
</table>

$\Sigma x n_x = 10094; \bar{x} = 3.870; \Sigma x^2 n_x = 486.50; S^2 = 3.676; n = 77.$

<table>
<thead>
<tr>
<th>n</th>
<th>77</th>
<th>79</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N[\log n - \log (n-x)]$</td>
<td>134.4390</td>
<td>130.9693</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>-0.1395</td>
<td>-0.0872</td>
</tr>
</tbody>
</table>

Moment estimates. We have for this data,

$N = 2608$

$\bar{x} = \Sigma x n_x = 3.870$

$S^2 = \Sigma (x-\bar{x})^2 n_x = 3.676$

so that the estimate for the mean number $\mu$ of $\alpha$-particles emitted per interval is $\hat{\mu} = \bar{x} = 3.870$,

and the number $(n)$ of particles capable of disintegration for the substance during the interval of 7.5 seconds is estimated by

$n = \frac{\bar{x}^2}{\bar{x}^2 - S^2} = 77.$

Maximum likelihood estimates. The estimate for the mean number of $\alpha$-particles per interval remains the same, namely $\hat{\mu} = 3.870$. To get the estimate $\hat{n}$ of $n$, starting with the moment estimate $n = 77$, we solve the equation:

$\Psi(n) = \Sigma \frac{T_{x+1}}{n-x} - N[\log n - \log (n-x)] = 0,$
For \( n = 77 \), we have \( N \{ \log n - \log (n-x) \} = 134.4390 \). From column 4 of Table 1 we have for \( n = 77 \),

\[
\sum_{r > 0} \frac{T_{r+1}}{n - r} = \sum_{s > 1} \frac{T_s}{n-s+1} = 134.2995
\]

\[
\therefore \quad \Psi(77) = 134.2995 - 134.4390 = -0.1396.
\]

Let us try next \( n = 79 \), say. Now, \( N \{ \log n - \log (n-x) \} = 130.9693 \) and column 5 of the above table gives for \( n = 79 \),

\[
\sum_{r > 0} \frac{T_{r+1}}{n - r} = 131.0065
\]

\[
\therefore \quad \Psi(79) = 0.0372.
\]

Thus, whereas \( \Psi(77) \) is negative, \( \Psi(79) \) is positive and therefore the likelihood estimate for \( n \) is \( n > 77 \) and \( < 79 \).

\[
\therefore \quad \hat{n} = 78.
\]

**Example 2:** The first two columns of the following table give data, due to Weldon, that show the results of throwing \( n \) dice 4096 times, a throw of 4, 5 or 6 being called a success. \( x \) denotes the number of successes and \( n_x \) the frequency of \( x \).

\[
\text{TABLE 2}.
\]

<table>
<thead>
<tr>
<th>( x )</th>
<th>frequency ( n_x )</th>
<th>( T_x = \sum n_x ) ( s \geq x )</th>
<th>( \frac{T_x}{n-x+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>4096</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>4066</td>
<td>371.7275</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>4099</td>
<td>371.7275</td>
</tr>
<tr>
<td>3</td>
<td>198</td>
<td>4039</td>
<td>402.0000</td>
</tr>
<tr>
<td>4</td>
<td>430</td>
<td>3821</td>
<td>425.6667</td>
</tr>
<tr>
<td>5</td>
<td>731</td>
<td>3401</td>
<td>425.1250</td>
</tr>
<tr>
<td>6</td>
<td>948</td>
<td>2670</td>
<td>381.4285</td>
</tr>
<tr>
<td>7</td>
<td>847</td>
<td>1722</td>
<td>287.0000</td>
</tr>
<tr>
<td>8</td>
<td>536</td>
<td>875</td>
<td>175.0000</td>
</tr>
<tr>
<td>9</td>
<td>257</td>
<td>339</td>
<td>84.7600</td>
</tr>
<tr>
<td>10</td>
<td>71</td>
<td>82</td>
<td>27.3333</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>11</td>
<td>5.6000</td>
</tr>
<tr>
<td>total</td>
<td>4096</td>
<td>3972.7641</td>
<td>2600.6383</td>
</tr>
</tbody>
</table>
GENERALIZED POWER SERIES DISTRIBUTION

\[ \Sigma \epsilon n_a = 25145; \bar{x} = 6.139; \Sigma \epsilon^2 n_a = 166367; S^2 = 2.930; n = 12. \]

<table>
<thead>
<tr>
<th>n</th>
<th>12</th>
<th>13</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N[\log n - \log (n-x)])</td>
<td>2935.1030</td>
<td>2617.7000</td>
<td>3344.9000</td>
</tr>
<tr>
<td>(\Psi)</td>
<td>-7.3389</td>
<td>-17.0617</td>
<td>+21.9123</td>
</tr>
</tbody>
</table>

**Moment estimates.** We have

\[ N = 4096 \]
\[ \bar{x} = 6.139 \]
\[ S^2 = 2.930 \]

so that the estimate for the number of dice thrown is given by

\[ n = \frac{x^2}{\bar{x} - s^2} = 12, \]

and the estimate of the proportion of successes \(\pi\) is

\[ \pi = \frac{x}{n} = \frac{6.139}{12} = 0.5116. \]

**Maximum likelihood estimates.** To get firstly the estimate \(\hat{n}\) of \(n\), starting with the moment-estimate \(n = 12\), we solve the equation:

\[ \Psi(n) = \sum \frac{T_{r+1}^s}{n-r} - N[\log n - \log (n-x)] = 0. \]

For \(n = 12\), we have \(N[\log n - \log (n-x)] = 2935.1030\). From column 4 of Table 2 we have for \(n = 12\),

\[ \sum \frac{T_{r+1}^s}{n-r} = \sum \frac{T_{r+1}^s}{n-x+1} = 2927.7641 \]

\[ \therefore \psi(12) = 2927.7641 - 2935.1060 = -7.3380. \]

Let us try next \(n = 13\), say, to see if \(\psi(13)\) is near zero. We get by proceeding as before, \(\psi(13) = -17.0617\) which is further from zero than \(\psi(12)\). Therefore, we try \(n = 11\). We have then \(\psi(11) = 21.9123\) which indicates that \(\hat{n} = 12\). The estimate of \(\pi\) is then obtained by \(\hat{n} = \frac{x}{\hat{n}} = 0.5116\). *(Note: Weldon had thrown 12 dice).*
References

Bliss, C. I. (1953): Fitting the negative binomial distribution to biological data, with a note on the efficient fitting of the negative binomial distribution. *Biometrics*, 9, 178-200.


