THE NEGATIVE BINOMIAL DISTRIBUTION:

COMPUTATION OF THE MEDIAN AND THE MEAN ABSOLUTE DEVIATION

Definitions

A discrete random variable $x$ is called a discrete distribution when $n \geq 0$ numbers precede the geometric distribution with its probability density function $f(x; n, p)$.

Mean, Va

$M$ $V$

These values are computed using the probability density function $f(x; n, p)$.
THE NEGATIVE BINOMIAL DISTRIBUTION:
COMPUTATION OF THE MEDIAN AND THE MEAN ABSOLUTE DEVIATION

Definitions

For $n > 0$, $0 < p < 1$, and $q = 1 - p$, the distribution of the discrete variable $x$, having frequency function

$$f(x; n, p) = \binom{x+n-1}{x} p^n q^x = \binom{-n}{x} p^n (-q)^x, \ x = 0, 1, 2, \ldots,$$

is called the negative binomial distribution. It also is called Pascal's distribution when $n$ is a positive integer, and is called the geometric distribution when $n = 1$.

If $n$ is a positive integer there are two well-known instances of this distribution. In a sequence of Bernoulli trials with probability $p$ of success, $f(x; n, p)$ is the probability that the $n^{th}$ success will occur on the trial numbered $(n+x)$; that is, it is the probability that exactly $x$ failures will precede the $n^{th}$ success. Also, $f(x; n, p)$ is the frequency function of the distribution of the sum of $n$ random, independent variables, each of which has the geometric distribution with frequency function $f(x; 1, p)$; that is, $f(x; n, p)$ is the frequency function of the $n^{th}$ convolution of the geometric distribution with itself, which will be evident from the generating function.

Mean, Variance, and Higher Moments

Mean, $\mu = n q/p$

Variance from the mean, $\sigma^2 = n q/p^2$

These values can be obtained by direct summation or from the generating function.

From: James M. Dobbie
\[ F(s) = \sum_{x=0}^{\infty} f(x; n, p) \cdot s^x = \left( \frac{p}{1-q}s \right)^n = p^n(1-q) - n \]

Higher moments can be computed from the formula

\[ \sum_{x=k}^{\infty} \binom{x}{k} f(x; n, p) = \left( \binom{n+k-1}{k} \right) \left( \frac{p}{1-q} \right)^k \]

obtained by taking the \( k \) th derivative of \( F(s) \) and putting \( s = 1 \). Such moments are needed in fitting polynomials in exponential smoothing. Thus, if

\[ p_{t+t} = \sum_{(k)} a_k(T) \cdot t^k \]

the \( n \) th exponentially-weighted average of the values of the polynomial for \( t \leq 0 \) is

\[ S^n_T(p) = \alpha^n \sum_{x=0}^{\infty} \beta^x \binom{x+n-1}{n} p_{t-x} \]

\[ = \sum_{(k)} (-1)^k \binom{x}{k} \sum_{x=0}^{\infty} x^k f(x; n, \alpha) \]

However, if the polynomial is written in the form

\[ p_{t+t} = \sum_{(k)} b_k(T) \left( \frac{t+k-1}{k} \right) \]

the \( n \) th average is

\[ S^n_T(p) = \sum_{(k)} (-1)^k b_k(T) \sum_{x=k}^{\infty} \binom{x}{k} f(x; n, \alpha) \]

\[ = \sum_{(k)} (-1)^k \binom{n+k-1}{k} \left( \frac{p}{1-q} \right)^k b_k(T). \]

From: James M. Dobbie

Arthur D. Little Inc.
It is for this reason that the second form of the polynomial is the preferred form.

Median

In general, the equation

\[ \sum_{x=0}^{m} f(x;n,p) = 1/2 , \]

has no integral solution \( m \). However, just as in the positive binomial distribution, the partial sum can be written in terms of the Incomplete Beta-function, which then can be used as the definition of the partial sum for non-integral values of \( m \). In this way we can find a value of \( m \), not necessarily an integer, that satisfies the equation. This value of \( m \) will be called the median.

The formula for the positive binomial distribution is

\[ \sum_{x=0}^{m} \binom{n}{x} p^x q^{n-x} = I_q(n-m,m+1) = 1 - I_p(m+1,n-m), \]

(1)

where

\[ I_p(a,b) = \frac{\int_0^p x^{a-1}(1-x)^{b-1} \, dx}{\int_0^1 x^{a-1}(1-x)^{b-1} \, dx} \]

is the Incomplete Beta-function. The formula appears in many books and can be proved easily by integration by parts.

The corresponding formula for the negative binomial distribution is

\[ \sum_{x=0}^{m} \binom{x+n-1}{x} p^n q^x = I_p(n,m+1) = 1 - I_q(m+1,n). \]

(2)

From: James M. Dobbie

Arthur D. Little Inc.
This formula is not readily available. It is stated, but not prominently, in the Introduction to Pearson's Tables of the Incomplete Beta-function and Pearson gives a proof (provided by Fieller) in Biometrika, Vol. XXV, pp. 160-161.

A simpler proof is the following: Integrating by parts,

\[
\int_0^1 u^{n-1}(1-u)^m \, du = \frac{1}{n} \int_0^1 u^n q^m + \frac{m}{n} \int_0^1 u^n (1-u)^{m-1} \, du
\]

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

Hence

\[
I_p(n,m+1) = p^n \sum_{k=0}^m \binom{n+m}{k} p^{m-k} q^k
\]  \hspace{1cm} (3)

By induction on \( m \) it is easy to show that

\[
\sum_{k=0}^m \binom{n+m}{k} p^{m-k} q^k = \sum_{k=0}^m \binom{x+n-1}{x} q^x.
\]  \hspace{1cm} (4)

Formula (2) is obtained from (3) and (4).

Although formula (2) has a meaning only when \( m \) is a non-negative integer, the integrals in the Incomplete Beta-function exist for non-integral values of \( m \). We define the median of the negative binomial distribution to be the solution \( m \) of the equation

\[
I_p(n,m+1) = 1/2
\]  \hspace{1cm} (5)
A unique solution $m \geq 0$ exists, provided $p^n \leq 1/2$.

**Mean Absolute Deviation**

The mean absolute deviation from the median is

$$\Delta = \sum_{x=0}^{\infty} |x-m| f(x;n,p)$$

Let

$$[m] = \text{integral part of } m$$

$$\text{ } = \text{largest integer that does not exceed } m.$$  

Then

$$\Delta = \sum_{x=0}^{[m]} (m-x)f(x;n,p) + \sum_{m+1}^{\infty} (x-m)f(x;n,p)$$

$$= m \left[ 2 \sum_{1}^{[m]} f(x;n,p) - 1 \right] + m - 2 \sum_{1}^{[m]} x f(x;n,p)$$

Since

$$xf(x;n,p) = \mu f(x-1;n+1,p),$$

$$\Delta = m \left[ 2 \mu f(n, [m] +1) - 1 \right] + m \left[ 2 \mu f(n+1, [m]) \right]$$

Other forms for $\Delta$ can be obtained from

$$\sum_{1}^{[m]} f(x-1;n+1,p) = \sum_{1}^{[m]} f(x;n,p) - \frac{1}{\mu} f([m],n+1,p)$$

$$= \sum_{1}^{[m]} f(x;n,p) - \frac{([m] +1)}{\mu} f([m] +1;n,p)$$

From: James M. Dobbie

Arthur D. Little, Inc.
Two of these are

\[ \Delta = (\mu - m) \left[ \frac{2}{3} \Gamma_p(n, [m] + 1) \right] + 2 \mu \binom{n}{m} \]

and

\[ \Delta = (\mu - m) \left[ \frac{2}{3} \Gamma_p(n, [m] + 1) \right] + \frac{2}{p} \left[ \Gamma_p(n, [m] + 1) \right] \left[ \Gamma_p(n, [m] + 2) - \Gamma_p(n, [m] + 1) \right] \tag{6} \]

The latter form is easy to use, since

\[ [m] + 1 \leq m + 1 < [m] + 2 ; \]

that is, the two arguments involved are the two integers between which we inter-

polate in finding the solution of (5). Thus, to find \( \Delta \), enter the tables

of the Incomplete Beta-function and record the values

\[ \Gamma_p(n, [m] + 1) \text{ and } \Gamma_p(n, [m] + 2) \]

for which

\[ \Gamma_p(n, m + 1) < 1/2 \text{ and } \Gamma_p(n, [m] + 2) > 1/2 \]

when the second argument has integral values. Interpolate to find \( m \) for which

\[ \Gamma_p(n, m + 1) = 1/2 \]

and then substitute in (6).

If \( m \) is an integer,

\[ \Delta = \frac{m + 1}{p} \left[ 2 \Gamma_p(n, m + 2) - 1 \right] + 2n \binom{n}{m} p^{n-1} q^{m+1} \]

From: James M. Dobbie

Arthur & Little, Inc.
If \( m \) is not an integer and we use linear interpolation between integral values to find it, then

\[
\Delta = \left( \mu - m + \frac{1+ [m]}{p(m - [m])} \right) \left[ 1 - 2 I_p(n, [m] + 1) \right]
\]

A quantity of interest in inventory problems is the expected amount by which \( x \) exceeds a given value \( k \), that is, the expected back-order

\[
B = \sum_{x=k}^{\infty} (x-k) f(x; n, p).
\]

By the same arguments used to find \( \Delta \) we find

\[
B = (\mu - k) \left[ 1 - I_p(n, k+1) \right] + \frac{(1+k)}{p} \left[ I_p(n, k+2) - I_p(n, k+1) \right]
\]

for the negative binomial distribution.

**Examples**

The values of \( \Delta \) and \( \sigma \) listed in the table below were computed primarily to test the hypothesis that

\[
\Delta = k \sigma,
\]

where \( k \) is 0.75 approximately. For \( p = 0.1 \) it is necessary to use the formula

\[
I_p(a, b) = 1 - I_q(b, a)
\]

From: James M. Dobbie

Arthur D. Little Inc.
For example, $I_{0.9}(6,1) = 0.5314$ and $I_{0.9}(7,1) = 0.4783$; from which $I_{0.1}(1,6) = 0.4686$ and $I_{0.1}(1,7) = 0.5217$.

<table>
<thead>
<tr>
<th>p</th>
<th>n</th>
<th>$\mu$</th>
<th>$a$</th>
<th>$m$</th>
<th>$\mu - m$</th>
<th>$\Delta / \sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
<td>9</td>
<td>9.5</td>
<td>5.6</td>
<td>3.4</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>18</td>
<td>13.4</td>
<td>14.5</td>
<td>3.5</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>27</td>
<td>16.4</td>
<td>23.4</td>
<td>3.6</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>36</td>
<td>19.0</td>
<td>32.4</td>
<td>3.6</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>45</td>
<td>21.2</td>
<td>41.4</td>
<td>3.6</td>
<td>0.77</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.71</td>
<td>0</td>
<td>1</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>2.00</td>
<td>1</td>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.45</td>
<td>2</td>
<td>1</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>2.83</td>
<td>3</td>
<td>1</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>3.16</td>
<td>4</td>
<td>1</td>
<td>0.78</td>
</tr>
<tr>
<td>0.9</td>
<td>10</td>
<td>1.11</td>
<td>1.11</td>
<td>0.43</td>
<td>0.68</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.22</td>
<td>1.57</td>
<td>1.51</td>
<td>0.71</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>3.33</td>
<td>1.92</td>
<td>2.63</td>
<td>0.70</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>4.44</td>
<td>2.22</td>
<td>3.74</td>
<td>0.70</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>5.56</td>
<td>2.48</td>
<td>4.85</td>
<td>0.71</td>
<td>0.81</td>
</tr>
</tbody>
</table>