**Finding *E(Y)* and *Var(Y)* for some Discrete Probability Distributions**

**Dr Richard Kenderdine**

www.kenderdinemathstutoring.com.au

This note derives the formulae for the expected value and variance of a number of standard discrete probability distributions – binomial, geometric, hypergeometric, negative binomial and Poisson.

If $p(y)$ is a probability function then the expected value is

$$E(Y) = \sum_y yp(y)$$

and the variance is obtained from

$$\text{Var}(Y) = E(Y^2) - \big(E(Y)\big)^2$$

Sometimes it is more efficient to determine these expressions by using the Moment Generating Function.

**The Moment Generating Function**

The $k^{th}$ moment of a random variable $Y$ taken about the origin is defined to be $E(Y^k)$.

The moment generating function $m(t)$ for a random variable $Y$ is defined to be

$$m(t) = E(e^{tY})$$

The $k^{th}$ moment is then the $k^{th}$ derivative of $m(t)$ evaluated at $t = 0$.

## Binomial Probability Distribution

The binomial probability distribution is used for a random variable that has only two possible outcomes (termed as success or failure) with a fixed probability of occurring in each of a fixed number of trials. The outcome of each trial is independent of the outcome of every other trial.

The probability of $y$ successes in $n$ trials with probability of success $\pi$ is

$$P(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$$

This comes from the $\binom{n}{y}$ ways of obtaining $y$ successes from $n$ trials, each with a probability of $\pi^y(1-\pi)^{n-y}$.

**Expected value**

The expected value of $y$ is given by

$$E(Y) = \sum_{y=0}^{n} y \binom{n}{y} \pi^y (1-\pi)^{n-y} = \sum_{y=1}^{n} y\pi \left(\frac{n}{y}\right) \binom{n-1}{y-1} \pi^{y-1} (1-\pi)^{n-y}$$

$$= n\pi \sum_{x=o}^{n-1} \binom{n-1}{x} \pi^x (1-\pi)^{n-1-x} = n\pi$$

Here the substitution $x = y - 1$ is used and the summation over $x$ equals 1.

**Variance**

We use the fact that $y^2 = y(y-1) + y$ to find $E(Y^2) = E\big(Y(Y-1)\big) + E(Y)$

$$E(Y^2) = \sum_{y=2}^{n} y(y-1)\pi^2 \left(\frac{n(n-1)}{y(y-1)}\right) \binom{n-2}{y-2} \pi^{y-2} (1-\pi)^{n-y} + \eta\pi$$

$$= \pi^2 n(n-1) + n\pi$$

Then using $\mathrm{Var}(Y) = E(Y^2) - \big(E(Y)\big)^2$ we have

$$\mathrm{Var}(Y) = n^2\pi^2 - n\pi^2 + n\pi - (n\pi)^2 = n\pi(1-\pi)$$

**Moment generating function**

The mgf of the Binomial Probability Distribution is given by

$$m(t) = \sum_{y=0}^{n} e^{ty} \binom{n}{y} \pi^{y} (1-\pi)^{n-y}$$

Collecting terms with power $y$ we have

$$\sum_{y=0}^{n} \binom{n}{y} (1-\pi)^{n} \left(\frac{\pi e^{t}}{1-\pi}\right)^{y} = (1-\pi)^{n} \left(1 + \frac{\pi e^{t}}{1-\pi}\right)^{n} = (1-\pi+\pi e^{t})^{n}$$

Here use is made of the Binomial expansion

$$(1+x)^{n} = \sum\nolimits_{x=0}^{n} \binom{n}{x} x^{n}$$

Then $E(Y) = m'(0)$ with $m'(t) = n(1-\pi+\pi e^{t})^{n-1}\pi e^{t}$.

Then $m'(0) = n\pi = \ E(Y)$

And $E(Y^2) = m''(0)$ with $m''(t) = n(1-\pi+\pi e^{t})^{n-2}\pi e^{t}\big((n-1)\pi e^{t} + (1-\pi+\pi e^{t})\big)$

Then $m''(0) = n\pi\big((n-1)\pi + 1\big) = n^2\pi^2 - n\pi^2 + n\pi$

Hence $\mathrm{Var}(Y) = n^2\pi^2 - n\pi^2 + n\pi - (n\pi)^2 = n\pi(1-\pi)$

## Geometric Probability Distribution

The geometric random variable $Y$ is the number of the trial at which the first success occurs, with probability of success at any trial being $p$. Then

$$P(Y = k) = (1-p)^{k-1}p$$

**Calculation of *E(Y)***

Let $q = 1 - p$ then

$$E(Y) = \sum_{k=1}^{\infty} kq^{k-1}p = p + 2qp + 3q^2p + \cdots$$

$$= \frac{d}{dq}[qp + q^2p + q^3p + \cdots]$$

$$= p\frac{d}{dq}[q + q^2 + q^3 + \cdots]$$

Now $q + q^2 + q^3 + \cdots$ is an infinite geometric series with first term and ratio both $q$. Since $0 < q < 1$ we can sum this series using $S_\infty = \frac{a}{1-r} = \frac{q}{1-q}$. Therefore

$$E(Y) = p\frac{d}{dq}\left[\frac{q}{1-q}\right] = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

(using the quotient rule to differentiate).

Hence, we have $E(Y) = \frac{1}{P}$

**Calculation of *Var(Y)***

Since *Var(Y)* = $E(Y^2) - [E(Y)]^2$ we first need to calculate $E(Y^2)$

$$E(Y^2) = \sum_{k=1}^{\infty} k^2q^{k-1}p = p + 4qp + 9q^2p + \cdots$$

$$= p(1 + 4q + 9q^2 + \cdots\cdot)$$

$$= p\frac{d}{dq}(q + 2q^2 + 3q^3 + \cdots)$$

$$= p\frac{d}{dq}[(q + q^2 + q^3 + \cdots) + (q^2 + q^3 + \cdots) + (q^3 + q^4 + \cdots)]$$

Now we have an infinite series of infinite series, all with ratio $q$ but different first terms. Hence, we have inside the brackets, the sum of an infinite number of formulae for summing an infinite series.

$$E(Y^2) = p\frac{d}{dq}[\frac{q}{1-q} + \frac{q^2}{1-q} + \frac{q^3}{1-q} + \cdots]$$

$$= p\frac{d}{dq}\left[\frac{1}{1-q}(q + q^2 + q^3 + \cdots)\right]$$

$$= p\frac{d}{dq}\left[\frac{1}{1-q} \cdot \frac{q}{1-q}\right]$$

$$= p\frac{d}{dq}\left[\frac{q}{(1-q)^2}\right]$$

$$= p\left(\frac{1+q}{(1-q)^3}\right) = \frac{p(1+q)}{p^3} = \frac{1+q}{p^2} = \frac{1+(1-p)}{p^2} = \frac{2-p}{p^2}$$

(again, using the quotient rule for differentiating).

Finally,

$$\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

**Useful properties of the Geometric Distribution**

1) $P(Y > a) = q^a$

   Proof:

$$P(Y > a) = \sum_{k=0}^{\infty} q^{a+k}p = q^a p \sum_{k=0}^{\infty} q^k = q^a p\left(\frac{1}{1-q}\right) = q^a p\left(\frac{1}{p}\right) = q^a$$

2) $$P(Y > a + b | Y > a) = \frac{P(Y>a+b \ \cup Y>a)}{P(Y>a)} = \frac{P(Y>a+b)}{P(Y>a)} = \frac{q^{a+b}}{q^a} = q^b$$

This is the memoryless property that says if it is known that there are more than $a$ trials then effectively the distribution resets to zero and the probability depends only on the number of trials beyond $a$.

**Negative Binomial Probability Distribution**

In the geometric distribution the random variable is the trial on which the first success occurs. This is extended with the negative binomial random variable which is the trial on which the $r^{th}$ success occurs (*r* = 2, 3, ,...). That is, the number of trials varies for a fixed number of successes, contrasting with binomial probability that has a fixed number of trials and a varying number of successes.

If *p* is the probability of success on any trial then the probability that the $r^{th}$ success occurs on the $y^{th}$ trial ($y \geq r$) is

$$P(y) = \binom{y-1}{r-1} p^{r-1} q^{y-1-(r-1)} p = \binom{y-1}{r-1} p^r q^{y-r}$$

That is, there are $r - 1$ successes on the first $y - 1$ trials and then immediately followed by the $r^{th}$ success.

Calculation of *E(Y)* from $\sum yp(y)$ and *Var(Y)* = $\sum y^2\, p(y) - (E(Y))^2$ is complicated in this case and another method is presented.

**The Moment Generating Function.**

Fortunately, we can use the mgf for the negative binomial distribution.

$$m(t) = \sum_{y=r}^{\infty} e^{ty} \binom{y-1}{r-1} p^r q^{y-r}$$

$$= \left(\frac{p}{q}\right)^r \sum_{y=r}^{\infty} \binom{y-1}{r-1} (qe^t)^y$$

It is inconvenient to commence the summation at $y = r$ so we use the transformation $x = y - r$ and therefore $y = x + r$ to give

$$m(t) = \sum_{x=0}^{\infty} e^{t(x+r)} \binom{x+r-1}{r-1} p^r q^x$$

$$= (pe^t)^r \sum_{x=0}^{\infty} \binom{x+r-1}{r-1} (qe^t)^x$$

The summation looks a bit like the expansion of a binomial $(1 + x)^n$, where $n$ is a positive integer, except that the summation has an infinite number of terms and the binomial coefficient $\binom{x+r-1}{r-1}$ has a variable and increasing, as opposed to constant, first term. We can find what we want by considering the expansion of $(1-x)^n$ when $n$ is a negative integer.

**A diversion into infinite series**

Recall that the limiting sum of an infinite geometric series with first term 1 and ratio $x$, with $|x| < 1$, is given by $\frac{1}{1-x}$. That is, $\frac{1}{1-x} = 1 + x + x^2 + \cdots$ for $|x| < 1$.

Using this, we can differentiate both sides to give

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots$$

and again,

$$\frac{2}{(1-x)^3} = 2 + 6x + 12x^2 + \cdots$$

or

$$\frac{1}{(1-x)^3} = 1 + 3x + 6x^2 + \cdots$$

In general, we have

$$\frac{1}{(1-x)^n} = \sum_{r=0}^{\infty} \binom{n+r-1}{n-1} x^r$$

Note that $n$ is fixed and $r$ is varying. Compare this with the summation part of $m(t)$:

$$\sum_{x=0}^{\infty} \binom{x+r-1}{r-1} (qe^t)^x$$

where $r$ is fixed and $x$ is varying. Hence, we can simplify the moment function as

$$m(t) = (pe^t)^r \frac{1}{(1-qe^t)^r} = \frac{(pe^t)^r}{(1-(1-p)e^t)^r}$$

Since we need to differentiate this function it can be expressed more simply as

$$m(t) = \left[\frac{p}{e^{-t} - (1-p)}\right]^r$$

Or, even better as

$$m(t) = p^r[e^{-t} - (1-p)]^{-r}$$

Now find the first two derivatives (with respect to $t$):

$$m'(t) = p^r(-r)\left(e^{-t} - (1-p)\right)^{-r-1}(-e^{-t})$$

$$= rp^r(e^{-t})\left(e^{-t} - (1-p)\right)^{-r-1}$$

$$m''(t) = rp^r[-e^{-t}\left(e^{-t} - (1-p)\right)^{-r-1} - e^{-t}(r+1)\left(e^{-t} - (1-p)\right)^{-r-2}(-e^{-t})]$$

$$= rp^r[e^{-t}\left(e^{-t} - (1-p)\right)^{-r-2}][-\left(e^{-t} - (1-p)\right) + (r+1)e^{-t}]$$

$$= rp^r[e^{-t}\left(e^{-t} - (1-p)\right)^{-r-2}][1 - p + re^{-t}]$$

Now we can find the first two moments:

$$E(Y) = m'(0) = rp^r(1)\left(1 - (1-p)\right)^{-r-1} = \frac{r}{p}$$

$$E(Y^2) = m''(0) = rp^r\left[1\left(1 - (1-p)\right)^{-r-2}\right][1 - p + r(1)] = \frac{(1 + r - p)r}{p^2}$$

Hence

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{(1 + r - p)r}{p^2} - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}$$

Therefore, the negative binomial random variable has $E(Y) = \frac{r}{P}$ and $\text{Var}(Y) = \frac{r(1-p)}{p^2}$

**One use of the Negative Binomial Distribution**

The Poisson Distribution is used to model count data, that is, data that are frequencies. For example, arrivals to a queue per time period. Two disadvantages of the Poisson Distribution are that it is defined by only one parameter and the mean and variance are equal. Sometimes situations arise where more flexibility is required to adequately fit a distribution to data. The Negative Binomial Distribution may be an alternative as it is a two-parameter distribution with variance not equal to the mean.

As given previously, $E(Y) = \frac{r}{p}$ and $\text{Var}(Y) = \frac{r(1-p)}{p^2}$. Note that the variance can be written as $\frac{r}{p}\left(\frac{1-p}{p}\right) = \frac{r}{p}\left(\frac{q}{p}\right) = \frac{q}{p}E(Y)$ ie the variance is a multiple of the mean and equal to the mean when $p = 0.5$. It is larger when $p < 0.5$ and smaller when $p > 0.5$.

To illustrate the difference between Poisson and Negative Binomial Distributions three random samples were obtained from the latter with histograms plotted. Overlaid were the Negative Binomial Density Function (PDF) and a Poisson fit. In the first two cases the sample size was 10 000 while the third case had a smaller sample size of 100. In each case the number of successes was 20. Note that the random variable in this note is the total number of trials whereas in the simulations it is the number of failures. Hence the total number of trials is 20 more than the number of failures.

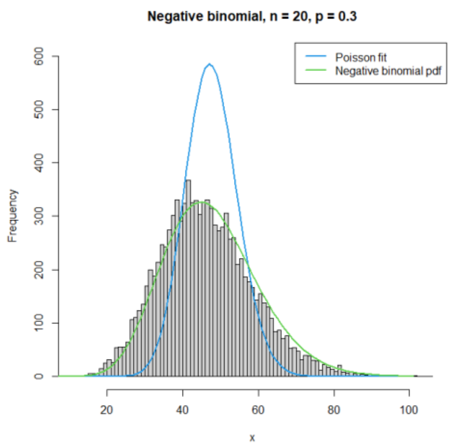


**Figure 1**: Sample size 10 000, number of successes 20 and probability of success 0.3

It can be seen from Figure 1 that the Poisson fit doesn't pick up the heavier tails exhibited in the Negative Binomial Distribution and as a result the former is higher in the centre. These characteristics are also shown in Figures 2 and 3, although perhaps the Poisson fit is good enough in the latter case.

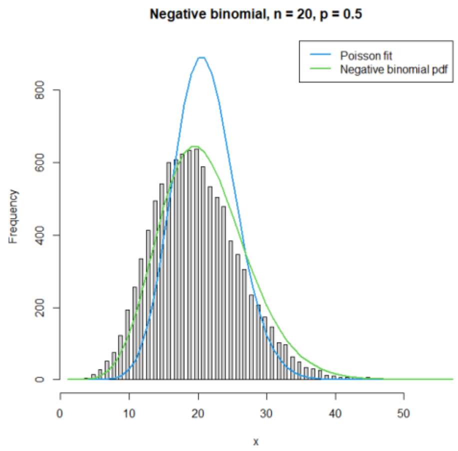


**Figure 2**: Sample size 10 000, number of successes 20 and probability of success 0.5

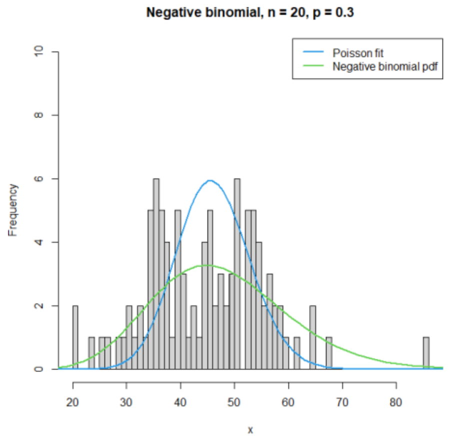


**Figure 3**: Sample size 100, number of successes 20 and probability of success 0.3

The R code for Figure 3 is

```
yrbnl003 = rnbinom(100, size = 20, prob = 0.3)
hist(yrbnl003, breaks=50, ylim = c(0,10),xlab = "x",main= "Negative binomial, n = 20, p = 0.3")
mean(yrbnl003)
45.12

lines(100*dpois(seq(0,100, by = 1), lambda=45.12),col = 4, lwd = 2)
lines(100*dnbinom(seq(0,100, by = 1), size = 20, prob = 0.3), col = 3,lwd = 2)
legend("topright", legend = c("Poisson fit","Negative binomial pdf"), col = c(4,3), lwd = c(2,2))
```

**Hypergeometric Probability Distribution**

.

The hypergeometric distribution is used when a sample of size $n$ is taken without replacement from a population of size $N$ with $r$ elements having a fixed characteristic and $N-r$ without the characteristic. For example, a Combinatorics question might ask to calculate the probability of a committee of 5 people selected from 8 men and 12 women containing exactly 2 men. The answer is

$$\frac{\binom{8}{2}\binom{12}{3}}{\binom{20}{5}}$$

Similarly, the probability of selecting a sample of size $n$, without replacement, with $y$ elements from $r$ and $n-y$ from $N-r$ is

$$P(y)=\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

This is the hypergeometric probability distribution of the random variable $Y$, the number of elements from $r$ in the sample.

**Use of Indicator Variables**

To facilitate the calculation of the expected value and variance of $Y$ we make use of the indicator variable $X$:

$$X_i=\begin{cases}1 & if\ element\ i\ has\ the\ characteristic\\ 0 & otherwise\end{cases}$$

Now $P(X_1=1)=\frac{r}{N}$ and to calculate $P(X_2=1)$ we need to account for $X_1$ to be both 0 and 1:

$$P(X_2=1)=P(X_1=0,X_2=1)+P(X_1=1,X_2=1)$$

$$=\frac{r}{N}\frac{r-1}{N-1}+\frac{N-r}{N}\frac{r}{N-1}=\frac{r}{N(N-1)}(r-1+N-r)=\frac{r}{N}$$

Similarly, for $P(X_3 = 1)$ we need to account for the outcomes of the first two draws:

| $X_1$ | $X_2$ | $X_3$ | Number of ways |
|---|---|---|---|
| 0 | 0 | 1 | $(N-r)(N-r-1)r$ |
| 0 | 1 | 1 | $(N-r)r(r-1)$ |
| 1 | 0 | 1 | $r(N-r)(r-1)$ |
| 1 | 1 | 1 | $r(r-1)(r-2)$ |

If we add the number of ways for the first two and factorise we have

$$r(N-r)(N-r-1+r-1) = r(N-r)(N-2)$$

Similarly for the last two

$$r(r-1)(N-r+r-2) = r(r-1)(N-2)$$

Now combine these results

$$r(N-2)(N-r+r-1) = r(N-1)(N-2)$$

Hence,

$$P(X_3 = 1) = \frac{r(N-1)(N-2)}{N(N-1)(N-2)} = \frac{r}{N}$$

In general, therefore

$$P(X_i = 1) = \frac{r}{N} \qquad \text{for } i = 1,2,\ldots,n$$

**Calculation of $E(Y)$ and $Var(Y)$**

Since $Y$ is the number of elements chosen from $r$ in the population and $X_i = 1$ if the $i^{\text{th}}$ draw is an element from that group, then

$$Y = \sum_{i=1}^{n} X_i$$

To find $E(Y)$ we need to find $E(X_i) = 1\left(\frac{r}{N}\right) + 0\left(1 - \frac{r}{N}\right) = \frac{r}{N}$

Hence,

$$E(Y) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{r}{N} = n\left(\frac{r}{N}\right)$$

(This uses the property that the expected value of a sum of random variables equals the sum of the expected values).

The calculation of the variance of a sum of random variables is more complicated as it needs to account for the covariance between the pairs of random variables.

$$\text{Var}(Y) = \sum_i \text{Var}\,(X_i) + 2\sum_i \quad \sum_j \quad \text{Cov}(X_i, X_j)$$

Now, $\text{Var}(X_i) = E(X_i^2) - \left(E(X_i)\right)^2$ and since $X_i^2 = 0$ or 1 then $E(X_i^2) = \frac{r}{N}$.

Hence,

$$\text{Var}(X_i) = \frac{r}{N} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(1 - \frac{r}{N}\right)$$

Covariance is calculated using

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

Note that $X_i X_j = 1$ only if both $X_i$ and $X_j$ are 1 and the probability of this is $\frac{r}{N}\frac{r-1}{N-1}$, independent of $i$ and $j$.

Hence,

$$\text{Cov}(X_i, X_j) = \frac{r}{N}\frac{r-1}{N-1} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(\frac{r-1}{N-1} - \frac{r}{N}\right)$$

$$= -\frac{r}{N}\left(\frac{N-r}{N(N-1)}\right)$$

$$= -\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{1}{N-1}\right)$$

Now in $n$ terms there are $\frac{n(n-1)}{2}$ ways of pairing them so putting this all together we have

$$\text{Var}(Y) = \sum_i \text{Var}\,(X_i) + 2\sum_i \quad \sum_j \quad \text{Cov}(X_i, X_j)$$

$$= n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right) + \frac{2n(n-1)}{2}\left(-\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{1}{N-1}\right)$$

$$= n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(1 - (n-1)\left(\frac{1}{N-1}\right)\right)$$

$$= n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Hence,

$$E(Y) = \frac{nr}{N} \quad \text{and} \quad \text{Var}(Y) = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

**Alternative method**

The moment generating function (mgf) only exists as a summation and not in closed form:

$$m(t) = \sum_{y=0}^{n} e^{ty} \binom{r}{y}\binom{N-r}{n-y} / \binom{N}{n}$$

The first moment is obtained from

$$m'(0) = \sum_{y=0}^{n} y \binom{r}{y}\binom{N-r}{n-y} / \binom{N}{n}$$

This is the same as finding $E(Y)$ from the usual definition so there is no advantage by considering the mgf. To evaluate this sum we manipulate the factorials:

$$\sum_{y=0}^{n} y \binom{r}{y}\binom{N-r}{n-y} \Big/ \binom{N}{n} = \sum_{y=0}^{n} y \binom{r}{y}\left(\frac{(r-1)!}{(y-1)!\,(r-y)!}\right)\binom{N-r}{n-y} \Big/ \binom{N}{n}$$

$$= r \sum_{y=1}^{n} \binom{r-1}{y-1}\binom{N-r}{n-y} / \binom{N}{n}\left(\frac{(N-1)!}{(n-1)!\,(N-n)!}\right)$$

$$= \left(\frac{nr}{N}\right) \sum_{y=1}^{n} \binom{r-1}{y-1}\binom{N-r}{n-y} / \binom{N-1}{n-1}$$

We need to change the limits of summation. Let $x = y - 1$ and therefore the existing limits $1 \leq y \leq n$ become $0 \leq x \leq n-1$:

$$E(X) = \left(\frac{nr}{N}\right) \sum_{x=0}^{n-1} \binom{r-1}{x}\binom{N-r}{n-(x+1)} / \binom{N-1}{n-1}$$

$$= \left(\frac{nr}{N}\right) \sum_{x=0}^{n-1} \binom{r-1}{x}\binom{N-r}{n-1-x} / \binom{N-1}{n-1}$$

Now we can use one version of the Chu-Vandermonde identity:

$$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n}$$

Hence, we have $E(X) = \frac{\binom{nr}{N}\binom{N-1}{n-1}}{\binom{N-1}{n-1}} = \frac{nr}{N}$

To find the variance we first need to obtain $E(Y^2)$ and to do this we write

$$y^2 = y^2 - y + y = y(y-1) + y$$

and then

$$E(Y^2) = \sum_{y=0}^{n} y(y-1)\binom{r}{y}\binom{N-r}{n-y} / \binom{N}{n} + \sum_{y=0}^{n} y\binom{r}{y}\binom{N-r}{n-y} / \binom{N}{n}$$

The second summation is, from above, $\frac{nr}{N}$ and to simplify the first summation we use a similar method as previously:

$$\sum_{y=0}^{n} y(y-1)\binom{r}{y}\binom{N-r}{n-y} \Big/ \binom{N}{n}$$

$$= \sum_{y=2}^{n} y(y-1)\left(\frac{r(r-1)}{y(y-1)}\right)\binom{r-2}{y-2}\binom{N-r}{n-y} / \left(\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}\right)$$

Simplifying, and letting $x = y - 2$, we have

$$\frac{nr(r-1)(n-1)}{N(N-1)} \sum_{x=0}^{n-2} \binom{r-2}{x}\binom{N-r}{n-2-x} / \binom{N-2}{n-2}$$

Again, using the Chu-Vandermonde identity, we have

$$E(Y^2) = \frac{nr(r-1)(n-1)}{N(N-1)} + \frac{nr}{N}$$

Hence

$$\mathrm{Var}(Y) = E(Y^2) - \big(E(Y)\big)^2 = \frac{nr(r-1)(n-1)}{N(N-1)} + \frac{nr}{N} - \left(\frac{nr}{N}\right)^2$$

$$= \frac{nr}{N}\left(\frac{nr - n - r + 1}{N-1} + 1 - \frac{nr}{N}\right)$$

$$= \frac{nr}{N}\left(\frac{Nnr - nN - rN + N + N^2 - N - nrN + nr}{N(N-1)}\right)$$

$$= \frac{nr}{N}\left(\frac{-nN - rN + N^2 + nr}{N(N-1)}\right) = \frac{nr}{N}\left(\frac{N(N-r) - n(N-r)}{N(N-1)}\right) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

**Poisson Probability Distribution**

The Poisson Probability Distribution is used for counts, that is, the theoretical number of items in a given time period.

The probability function is derived by taking the limit of the Binomial Probability Distribution. If $\pi$ is the probability of one arrival during a subinterval then let $\lambda$ be the arrival rate in $n$ subintervals. Taking the limit as $n \to \infty$ we have

$$\lim_{n\to\infty} \binom{n}{y} \pi^y (1-\pi)^{n-y} = \lim_{n\to\infty} \binom{n}{y} \left(\frac{\lambda}{n}\right)^y \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-y+1)}{y!} \left(\frac{\lambda^y}{n^y}\right) \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-y}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-y+1)}{n^y} \left(\frac{\lambda^y}{y!}\right) \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-y}$$

$$= 1 \times \frac{\lambda^y}{y!} \times e^{-y} \times 1 = \frac{e^{-y}\lambda^y}{y!}$$

Here use is made of $\lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n = e^x$. Hence, we have the Poisson Probability Function:

$$P(y) = \frac{e^{-\lambda}\lambda^y}{y!}$$

where $\lambda$ is the arrival rate (counts per time period). The assumptions are that items arrive independently of each other and the arrival rate is directly related to the length of the time interval.

**Expected value**

$$E(Y) = \sum_{y=0}^{\infty} y \frac{e^{-\lambda}\lambda^y}{y!} \equiv \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{(y-1)!}$$

$$= \lambda \sum_{y=1}^{\infty} \frac{e^{-\lambda}\lambda^{y-1}}{(y-1)!} = \lambda \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = \lambda(1) = \lambda$$

**Variance**

$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2 = E(Y(Y-1)+Y) - \left(E(Y)\right)^2$$

$$E(Y(Y-1)+Y) = \sum_{y=0}^{\infty} \frac{y(y-1)e^{-\lambda}\lambda^y}{y!} + \lambda$$

$$= \sum_{y=2}^{\infty} \frac{e^{-\lambda}\lambda^y}{(y-2)!} + \lambda = \lambda^2 \sum_{y=2}^{\infty} \frac{e^{-\lambda}\lambda^{y-2}}{(y-2)!} + \lambda = \lambda^2 \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} + \lambda$$

$$= \lambda^2 + \lambda$$

Hence $\mathrm{Var}(Y) = \lambda^2 + \lambda - (\lambda)^2 = \lambda$

**Using the Moment Generating Function**

The mgf is given by $E(e^{ty}) = \sum_{y=0}^{\infty} e^{ty}\left(\frac{e^{-\lambda}\lambda^y}{y!}\right) = \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda e^t)^y}{y!}$

$$=e^{-\lambda}\exp(\lambda e^t)$$

Here use is made of $e^x = \sum_{x=0}^{\infty} \frac{x^n}{n!}$

Then $m'(t) = e^{-\lambda}\lambda e^t \exp(\lambda e^t)$ and hence $E(Y) = m'(0) = e^{-\lambda}\lambda \exp(\lambda) = \lambda$ and $m''(t) = e^{-\lambda}\lambda e^t \exp(\lambda e^t) + e^{-\lambda}(\lambda e^t)^2 \exp(\lambda e^t)$, giving $E(Y^2) = m''(0) = \lambda + \lambda^2$.

Then, as before, $\mathrm{Var}(Y) = \lambda + \lambda^2 - (\lambda)^2 = \lambda$

**Sum of Poisson Probability Distributions**

Suppose $X$ and $Y$ are distributed as Poisson with parameters $\lambda_1$ and $\lambda_2$ respectively. Then the sum, $X+Y$, is distributed as Poisson with parameter $\lambda_1 + \lambda_2$.

*Proof*

$$P(X+Y=k) = \sum_{i=0}^{k} \frac{e^{-\lambda_1}\lambda_1^i}{i!} \frac{e^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!}$$

$$=e^{-(\lambda_1+\lambda_2)}\lambda_2^k \sum_{i=0}^{k} \frac{1}{i!(k-i)!}\left(\frac{\lambda_1}{\lambda_2}\right)^i = e^{-(\lambda_1+\lambda_2)}\frac{\lambda_2^k}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!}\left(\frac{\lambda_1}{\lambda_2}\right)^k$$

$$= e^{-(\lambda_1+\lambda_2)}\frac{\lambda_2^k}{k!}\left(1+\frac{\lambda_1}{\lambda_2}\right)^k = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}$$

which is a Poisson Distribution with parameter $\lambda, +\lambda_2$. Note that the Binomial expansion has been used to reduce the summation to a closed form.

## Summary

| Distribution | Probability | Mean | Variance | MGF |
|---|---|---|---|---|
| Binomial | $\binom{n}{y} p^y (1-p)^{n-y}$ | $np$ | $np(1-p)$ | $[pe^t + (1-p)]^n$ |
| Geometric | $p(1-p)^{y-1}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ | $\frac{pe^t}{1-(1-p)e^t}$ |
| Hypergeometric | $\binom{r}{y}\binom{N-r}{n-y} / \binom{N}{n}$ | $\frac{nr}{N}$ | $n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$ | NA |
| Poisson | $\frac{\lambda^y e^{-\lambda}}{y!}$ | $\lambda$ | $\lambda$ | $\exp[\lambda(e^t - 1)]$ |
| Negative Binomial | $\binom{y-1}{r-1} p^r (1-p)^{y-r}$ | $\frac{r}{p}$ | $\frac{r(1-p)}{p^2}$ | $\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$ |