Poisson distribution:

A random variable is said to have a Poisson distribution if the probability mass function of *X* is:

$$p(x;\lambda)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x=0,1,2,\ldots \quad , \quad \lambda>0$$

It is easy to see that f(x) enjoys the properties of a *pdf* because clearly $f_X(x)>0$ when $x\in S$ and $f_X(x)=0$ when $x\notin S$.

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

Where

$$e^{\lambda}=1+\lambda+\frac{\lambda^{2}}{2!}+\frac{\lambda^{3}}{3!}+\ldots=\sum_{x=0}^{\infty}\frac{\lambda^{x}}{x!}$$

is the Taylor series expansion of e^{λ}.

The mean of the Poisson distribution is

$$\mu=E(X)=\sum_{x}xf_X(x)=\sum_{x=0}^{\infty}x\frac{e^{-\lambda}\lambda^{x}}{x!}=\sum_{x=1}^{\infty}x\frac{e^{-\lambda}\lambda^{x}}{x!}$$

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

To find the variance

$Var(X)=E(X^2)-[E(X)]^2=E(X^2)-E(X)+E(X)-[E(X)]^2$

$\qquad\qquad=E(X(X-1))+E(X)-[E(X)]^2$

$$E[X(X-1)]=\sum_{x=0}^{\infty}x(x-1)\frac{e^{-\lambda}\lambda^{x}}{x!}=e^{-\lambda}\sum_{x=2}^{\infty}\lambda^{2}\frac{\lambda^{x-2}}{(x-2)!}$$

$$=\lambda^{2}e^{-\lambda}e^{\lambda}=\lambda^{2}$$

$$E(X^2)=E[X(X-1)]+E(X)=\lambda^2+\lambda$$

$$\therefore \operatorname{var}(X)=E(X^2)-\mu^2=\lambda^2+\lambda-\lambda^2=\lambda$$

Thus for the Poisson distribution: $\mu=\sigma^2=\lambda$

The moment generating function of X is

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

From the series representation of the exponential function, we have that

$$M_X(t) = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

for all real values of t.
Hence, the characteristic function is:

$$\phi_X(t) = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it} - 1)}$$

Now

$$M'_X(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

and

$$M''_X(t) = (\lambda e^t)^2 e^{\lambda(e^t - 1)} + \lambda e^t e^{\lambda(e^t - 1)}$$

Thus we can obtain

$$\mu = M'_X(0) = \lambda$$

$$\sigma^2 = M''_X(0) - [M'(0)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Poisson distribution can be used to approximate probabilities for a binomial distribution.

Theorem:

Let the random variable X has a binomial distribution with parameters n and p, then for each value $x = 0, 1, 2, \ldots,$ and as $p \to 0$ with $np = \lambda$ constant.

$$\lim_{n\to\infty} C_x^n p^x (1-p)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}$$

Proof:

$$f_X(x) = C_x^n p^x (1-p)^{n-x}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)(n-x)!}{x!\,(n-x)!} P^x (1-p)^{n-x}$$

$$f_X(x) = \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right)\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

The result follows by taking the limit, and recalling from calculus that

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right) = 1$$

Also

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1 \quad \text{and} \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

It follows that for every fixed positive integer *x*,

$$f_X(x) \approx \frac{e^{-\lambda}\lambda^x}{x!}$$

And for $x = 0$, $f_X(0) = (1-p)^n = \left(1-\frac{\lambda}{n}\right)^n$

This approximation is reasonably good when $n \to \infty$, $p \to 0$ and $np=\lambda$ remains constant. In practice, the approximation is good if $n \geq 20$ and $p \leq 0.05$, and it is very good if $n \geq 100$ and $np \leq 10$.

Example 1:

Let X have a Poisson distribution with a mean of $\lambda = 5$. Then

$$p(X=3) = \frac{e^{-5}5^3}{3!} = p(X \leq 3) - p(X \leq 2) = 0.140$$

Example 2:

For a certain manufacturing industry the number of industrial accidents averages *3* per week.

a. Find the probability that no accidents will occur in a given week.

b. Find the probability that two accidents will occur in a given week.

$$p(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0,1,2,\dots \quad , \quad \lambda > 0$$

$$a.\ p(X=0) = \frac{e^{-3}3^0}{0!} = e^{-3} = 0.05$$

$$b.\ p(X=2) = \frac{e^{-3}3^2}{2!} = \frac{9}{2}e^{-3} = 0.225$$

Also

$$p(X \leq 4) = \sum_{x=0}^{4} \frac{e^{-3}3^x}{x!} = 0.815$$

And

$$p(X \geq 4) = 1 - p(X \leq 3) = 1 - 0.647 = 0.353$$

And

$p(X = 4) = p(X \leq 4) - p(X \leq 3) = 0.815 - 0.647 = 0.168$

In actual practice, Poisson probabilities are seldom obtained by direct substitution. Instead, we refer to tables of Poisson probabilities; also, we refer to suitable computer software.

Example 3:

Ten percent of the tools produced in a certain manufacturing process turn out to be defective.

a. Find the probability that in a sample of 10 tools chosen at random, exactly two will be defective.

b. What is the expected number of defective tools?

c. Use the Poisson approximation to find the probability that the number of defective tools is less than 3.

$\lambda = np = 1$

a. $P(X=2) = C_2^{10}(0.1)^2(0.9)^8 = 0.1937$

b. $E(X) = np = 10(0.1) = 1$

c. $P(X < 3) = P(X=0) + P(X=1) + P(X=2) = e^{-1}[1+1+1/2] = 5/2e^{-1} = 0.92$

Example 4:

Records show that the probability is 0.00005 that a car will have a flat tire while crossing a certain bridge. Use the Poisson distribution to approximate the binomial probabilities that among 10000 cars crossing the bridge:

a. Exactly two will have a flat tire.

b. At most two will have a flat tire.

$\lambda = np = 10000(0.00005) = 0.5$

a. $P(X = 2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{(0.5)^2 e^{-0.5}}{2!} = 0.0758$

b. $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$

$\qquad = 0.6065 + 0.3033 + 0.0758 = 0.9856$

The Poisson distribution provides an adequate representation for the following situations:

1. Number of telephone calls arriving at an exchange per unit time.
2. Number of α- particles emitted by a radioactive substance.
3. Number of deaths occurring due to, say, heart disease in a city having large population.
4. Number of typing errors per page in a big text.
5. Number of defects occurring in the long length of cloth being manufactured in a factory.

Hypergeometric distribution:

Suppose that a lot consists of N items, of which k are of one type (called success) and N-k are of another type (called failure). Suppose that n items are sampled randomly and sequentially from the lot, without replacement. Let $Y_i = 1$ if the ith draw results in a success and $Y_i = 0$ otherwise, $i = 1 \dots n$. And let X denote the total number of successes among the n sampled items, then,

$$f_X(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0, 1, 2, \dots \text{ for } x \le k, \text{ and } \quad n\text{-}x \le N\text{-}k$$

This formula is known as the hyper geometric probability distribution. It arises from a situation quite similar to the binomial, except that the trials are dependent.

To show that $f_X(x)$ is a pdf, we have

$$\sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{x=0}^{n} \binom{k}{x}\binom{N-k}{n-x} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1$$

Using the formula:

$$\sum_{i=1}^{m} \binom{a}{i}\binom{b}{m-i} = \binom{a+b}{m}$$

The mean of the hyper geometric distribution is as follows:

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \sum_{x=0}^{n} \frac{x \, \frac{k!}{x!\,(k-x)!}\binom{N-k}{n-x}}{\frac{N!}{n!\,(N-n)!}}$$

$$= \frac{nk}{N}\sum_{x=1}^{n} \frac{\frac{(k-1)!}{(x-1)!\,(k-x)!}\binom{N-k}{n-x}}{\frac{(N-1)!}{(n-1)!\,(N-n)!}} = \frac{nk}{N}\sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N-1}{n-1}}$$

Let $y = x\text{-}1$, $m = k\text{-}1$, $R = N\text{-}1$, $r = n\text{-}1$, then

$$E(X) = \frac{nk}{N}\sum_{y=0}^{r} \frac{\binom{m}{y}\binom{R-m}{r-y}}{\binom{R}{r}} = \frac{nk}{N} = np$$

The variance of the distribution is:

$Var(X) = E(X^2) - [E(X)]^2 = E(X(X-1)) + E(X) - [E(X)]^2$

$$E(X(X-1) = = \sum_{x=0}^{n} x(x-1) \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=0}^{n} \frac{x(x-1) \, \frac{k!}{x!\,(k-x)!}\binom{N-k}{n-x}}{\frac{N!}{n!\,(N-n)!}}$$

$$= \sum_{x=2}^{n} \frac{\frac{k(k-1)(k-2)!}{(x-2)!\,(k-x)!}\binom{N-k}{n-x}}{\frac{N(N-1)(N-2)!}{n(n-1)(n-2)!\,(N-n)!}}$$

$$= \frac{k(k-1)n(n-1)}{N(N-1)}\sum_{x=2}^{n} \frac{\frac{(k-2)!}{(x-2)!\,(k-x)!}\binom{N-k}{n-x}}{\frac{(N-2)!}{(n-2)!\,(N-n)!}}$$

Let $y = x\text{-}2$, $m = k\text{-}2$, $R = N\text{-}2$, $r = n\text{-}2$, then

$$E(X(X-1)) = \frac{k(k-1)n(n-1)}{N(N-1)} \sum_{y=0}^{r} \frac{\frac{(m)!}{(y)!\,(m-y)!}\binom{R-m}{r-y}}{\frac{(R)!}{(r)!\,(R-r)!}}$$

$$= \frac{k(k-1)n(n-1)}{N(N-1)} \sum_{y=0}^{r} \frac{\binom{m}{y}\binom{R-m}{r-y}}{\binom{R}{r}} = \frac{k(k-1)n(n-1)}{N(N-1)}$$

$$Var(x) = \frac{k(k-1)n(n-1)}{N(N-1)} + np - n^2p^2$$

$$= np[\frac{(k-1)(n-1)}{N-1} + 1 - n\frac{k}{N}]$$

$$= np[\frac{N(k-1)(n-1) + N(N-1) - kn(N-1)}{N(N-1)}]$$

$$= \frac{\text{np}}{\text{N(N}-1)}[Nkn - Nk - Nn + N + N^2 - N - Nkn + nk]$$

$$= \frac{np}{N(N-1)}[N(N-n) - k(N-n)] = np\frac{(N-n)}{(N-1)}\left(\frac{N-k}{N}\right)$$

$$= np\left(\frac{N-n}{N-1}\right)\left(1-\frac{k}{N}\right) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

$$\therefore Var(x) = npq\left(\frac{N-n}{N-1}\right)$$

Theorem

If *X~ Hyp(n, k, N)*, then for each value $x = 0, 1, 2, \ldots, n$ as $n \to \infty$ and $k \to \infty$ with $k/N \to p$, a positive constant,

$$\lim_{N\to\infty} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

This is intuitively reasonable since the binomial distribution is applicable when we sample with replacement, while the hyper geometric distribution is applicable when we sample without replacement. If the size of the collection sampled from is large, it should not make a great deal of difference whether or not a particular item is returned to the collection before the next one is selected.

Example1

As part of an air pollution survey, an inspector decides to examine the exhaust of six of a company's 24 trucks. If 4 of the company's trucks emit excessive amount of pollutants. What is the probability that none of them will be included in the inspector's sample?

x = 0, n = 6, k = 4, N = 24

$$P(X = 0) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{4}{0}\binom{20}{6}}{\binom{24}{6}} = 0.2880$$

E(X) = n(k/N) = 6(4/24) = 1

$$V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) = n\frac{k}{N}\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$= 6\left(\frac{4}{24}\right)\left(1-\frac{4}{24}\right)\left(\frac{24-6}{24-1}\right) = \frac{15}{23}$$

Example2:

Among the twenty applicants for a job, only twelve are actually qualified. If five of the applicants are randomly selected for an interview, find the probability that only two of the five will be qualified for the job.

N = 20, n = 5, x = 2, k = 12

$$P(X = 2) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = 0.238$$

Example3:

A lot consists of 100 fuses, suppose that the lot contains 20 defective fuses. If five fuses are selected at random without replacement. What is the probability that:

1. At least three def. fuses will be obtained.
2. Between 2 to 3 def. fuses will be obtained.
3. Find the mean and the variance of the obtained defective fuses.

N = 100, k = 20, n = 5

1.

$$p(X \geq 3) = \sum_{x=0}^{n} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \sum_{x=3}^{5} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}$$

or

$$p(X \geq 3) = 1 - p(X < 3) = 1 - p(X \leq 2)$$

$$= 1 - \sum_{x=0}^{2} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}$$

2.

$$p(2 \leq X \leq 3) = p(X = 2) + p(X = 3)$$

$$= \frac{\binom{20}{2}\binom{80}{3}}{\binom{100}{5}} + \frac{\binom{20}{3}\binom{80}{2}}{\binom{100}{5}}$$

3.

E(X) = n(k/N) = 5(20/100) = 1

4.

$$V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) = n\frac{k}{N}\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$
$$= 5\left(\frac{20}{100}\right)\left(1-\frac{20}{100}\right)\left(\frac{100-5}{100-1}\right) = \frac{76}{99}$$