Chapter II
Special Distributions

1 Binomial Distribution (James Bernoulli)

Consider a random experiment, which is so defined that it has only two possible outcomes which we call *success* and *failure*. Let p be the probability of a success and $q = 1\text{-}p$ be the probability of failure. Let the experiment be repeated n times. Also assume that the trails are independent and the probability of success p remains unaltered by trail to trail. Then obviously the number of success in n trails is a random variable. Let X denote the number of successes in n independent repetitions of this experiment and let $f(x)$ be the p.d.f. of X.

If the x repetitions in which success occurs are specified, the probability of success in these x repetitions and failures in the remaining $n - x$ repetitions is $p^x q^{n-x}$ by the multiplication theorem the repetitions being independent. The x trials in which success occurs may be specified in $\binom{n}{x}$ mutually exclusive ways. So the event $X = x$ can occur in $\binom{n}{x}$ mutually exclusive ways and probability of each is $p^x q^{n-x}$. So by addition theorem, the required probability is $\binom{n}{x} p^x q^{n-x}$.

i.e.

$$f(x) = \binom{n}{x} p^x q^{n-x},\ x = 0,1,\ \ldots,\ n,\ 0 \le p \le 1.$$

This distribution is called binomial distribution.

Definition:- A discrete random variable X is said to follow a binomial distribution with parameters n and p if its probability density function is given by

$$f(x) = \binom{n}{x} p^x q^{n-x},\ x = 0,1,\ \ldots,\ n,\ 0 \le p \le 1.$$

Properties

1. Moments.

Mean = $\mu_1' = E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$= \sum_{x=0}^{n} \frac{n!}{x-1!(n-x)!} p^x q^{n-x} = np \sum_{x=0}^{n} \frac{n-1!}{x-1!(n-x)!} p^{x-1} q^{n-x}$

$= np\ (p+q)^{n-1} = np.$

$\mu_2' = E(X^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} x^2 \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$= \sum_{x=0}^{n} [x(x-1)+x] \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$= \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + np \sum_{x=0}^{n} \frac{x\, n!}{x!(n-x)!} p^{x-1} q^{n-x}$

$= n(n-1)p^2 \sum_{x=0}^{n} \frac{n-2!}{x-2!(n-x)!} p^{x-2} q^{n-x} + np \sum_{x=0}^{n} \frac{n-1!}{x-1!(n-x)!} p^{x-1} q^{n-x}$

$= n(n-1)p^2 (p+q)^{n-2} + np\ (p+q)^{n-1} = n(n-1)p^2 + np.$

$V(X) = \mu_2 = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - (np)^2 = npq.$

Now we can have

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

$x^3 = x(x-1)(x-2) + 3x^2 - 2x$

$x^4 = x(x-1)(x-2)(x-3) + 6x^3 - 11x^2 + 6x$

$\mu_3' = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$

$\mu_4' = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$

From the above we get

$$\mu_3 = npq(q-p)$$

and

$$\mu_4 = npq[1 + 3(n-2)pq$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}.$$

2. Moment generating function

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (q+pe^t)^n.$$

3. Mode

Mode of the distribution is that value of the random variable for which the pdf is maximum. Let M be the mode then

$$f(M) \geq f(M+1) \text{ and } f(M) \geq f(M-1)$$

$$\binom{n}{M} p^M q^{n-M} \geq \binom{n}{M+1} p^{M+1} q^{n-M-1} \qquad (1)$$

and

$$\binom{n}{M} p^M q^{n-M} \geq \binom{n}{M-1} p^{M-1} q^{n-M+1} \qquad (2)$$

Simplifying (1) we get

$$\frac{n!}{M!\,(n-M)!} p^M q^{n-M} \geq \frac{n!}{(M+1)!\,(n-M-1)!} p^{M+1} q^{n-M-1}$$

$$\frac{q}{n-M} \geq \frac{p}{M+1}$$

$$Mq + q \geq np - Mp$$

$$M(p+q) \geq np - q$$

$$M \geq np - q \qquad (3)$$

Simplifying (2) we get

$$M \leq np + p \qquad (4)$$

From (3) and (4) we get

$$np - q \leq M \leq np + p$$

Substituting for q=1-p and simplifying we get

$$(n+1)p - 1 \leq M \leq (n+1)p$$

If $(n+1)p$ is integer the binomial distribution has two mode one at $(n+1)p-1$ and the other at $(n+1)p$. If $(n+1)p$ is not an integer then the integer part of $(n+1)p$ is the mode.

4. If X_i (i=1, 2, …, k) are independent binomial random variate with parameters (n_i, p). Then $S_k = \sum_{i=1}^{k} X_i$ has a binomial distribution with parameters $(\sum_{i=1}^{k} n_i,\ p)$. (This property is called additive or reproductive property).

   We have

$$M_{x_i}(t) = (q+pe^t)^{n_i}$$

   So

$$M_{s_k}(t) = \prod_{i=1}^{k} M_{x_i}(t) = \prod_{i=1}^{k}(q+pe^t)^{n_i} = (q+pe^t)^{\sum n_i}$$

   Which is the mgf of $B(\sum_{i=1}^{k} n_i,\ p)$. Hence $S_k = \sum_{i=1}^{k} X_i$ has a binomial distribution with parameters $(\sum_{i=1}^{k} n_i,\ p)$.

5. If $p = q = ½$ the distribution is symmetrical and when $p \neq q$ the distribution is a skewed distribution.
6. If n is very large and if neither p and q is too close to zero binomial distribution may be approximated by normal distribution.

Result

Let X be a binomial random variable B(n,p) and μ_r denote the rth central moment. Then

$$\mu_{r+1} = pq\left[\frac{d\mu_r}{dp} + nr\mu_{r-1}\right]$$

Proof

We have E(X) = np and

$$\mu_r = E(X-np)^r = \sum_{x=0}^{n}(x-np)^r \binom{n}{x} p^x(1-p)^{n-x}$$

So

$$\frac{d\mu_r}{dp} = \sum \binom{n}{x} [xp^{x-1}(1-p)^{n-x}(x-np)^r - (n-x)p^x(1-p)^{n-x-1}(x-np)^r$$

$$-nrp^x(1-p)^{n-x}(x-np)^{r-1}]$$

$$= \sum \binom{n}{x} (x-np)^{r+1} \binom{n}{x} p^{x-1}(1-p)^{n-x-1} - \sum \binom{n}{x} (x-np)^{r-1} \binom{n}{x} p^x(1-p)^{n-x}$$

$=\frac{1}{pq}\sum\binom{n}{x}(x-np)^{r+1}\binom{n}{x}p^xq^x$- $\sum\binom{n}{x}(x-np)^{r-1}\binom{n}{x}p^xq^x$.

$$=\frac{1}{pq}\mu_{r+1}-nr\mu_{r-1}.$$

Hence

$$\mu_{r+1}=pq\left[\frac{d\mu_r}{dp}+nr\mu_{r-1}\right].$$

2 Poisson distribution (Simeon Denis Poisson)

Suppose that events that occur over a period of time or space satisfy the following:

1. The numbers of events occurring in disjoint intervals of time are independent.
2. The probability that exactly one event occurs in a small interval of time Δ is $\Delta\lambda$, where , $\lambda > 0$:
3. It is almost unlikely that two or more events occur in a sufficiently small interval of time.
4. The probability of observing a certain number of events in a time interval Δ depends only on the length of Δ and not on the beginning of the time interval.

Let X denote the number of events in a unit interval of time or in a unit distance. Then, X is called the Poisson random variable with mean number of events λ, in a unit interval of time.

Definition:-

A discrete random variable X is said to follow a Poisson distribution with parameters λ if its probability density function is given by

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad x=0,1,\ldots,\ \lambda > 0.$$

The following are some examples of variates which are found to follow the Poisson distribution

1. The number of defective articles in lot supplied by a good company

2. Number of printing mistakes on each page of a book published by a good publisher.
3. Number of wrong numbers telephone calls received in an office during working hours per week.
4. The number of cases of a disease in different towns;
5. The number of mutations in given regions of a chromosome;
6. The number of particles emitted by a radioactive source in a given time;
7. The number of births per hour during a given day.
8. In such situations we are often interested in whether the events occur randomly in time or space, or not.
9. The number of automobile accidents per day on an interstate road;

Properties

1 **Moments.**

Mean = μ_1' = E(X) = $\sum_{x=0}^{\infty} x\,\frac{\lambda^x}{x!}\,e^{-\lambda} = \sum_{x=0}^{\infty}\frac{\lambda^x}{x-1!}\,e^{-\lambda} = \lambda e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{x-1}}{x-1!}$

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

μ_2' = E(X^2) = $\sum_{x=0}^{\infty} x^2\,\frac{\lambda^x}{x!}\,e^{-\lambda} = \sum_{x=0}^{\infty}[x(x-1)+x]\frac{\lambda^x}{x-1!}\,e^{-\lambda}$

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

V(X) = μ_2 = E(X^2) –[E(X)]^2 = $\lambda^2+\lambda-\lambda^2 = \lambda.$

Now we can have

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

$x^3 = x(x-1)(x-2)+3x^2-2x$

$x^4 = x(x-1)(x-2)(x-3)+6x^3-11x^2+6x$

$\mu_3' = \lambda^3+3\lambda^2+\lambda$

$\mu_4' = \lambda^4+6\lambda^3+7\lambda^2+\lambda$

From the above we get

$$\mu_3 = \lambda$$

and

$$\mu_4 = 3\lambda^2 + \lambda$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{\lambda}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}.$$

2 Moment generating function

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

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

3 Mode

Mode of the distribution is that value of the random variable for which the pdf is maximum. Let M be the mode then

$$f(M) \geq f(M+1) \text{ and } f(M) \geq f(M-1)$$

$$e^{-\lambda} \frac{\lambda^M}{M!} \geq e^{-\lambda} \frac{\lambda^{M+1}}{(M+1)!} \qquad (1)$$

and

$$e^{-\lambda} \frac{\lambda^M}{M!} \geq e^{-\lambda} \frac{\lambda^{M-1}}{(M-1)!} \qquad (2)$$

Simplifying (1) we get

$$1 \geq \frac{\lambda}{(M+1)} \qquad (3)$$

Simplifying (2) we get

$$\frac{\lambda}{M} \leq 1 \qquad (4)$$

From (3) and (4) we get

$$\lambda - 1 \leq M \leq \lambda$$

If λ is integer the Poisson distribution has two mode one at λ -1 and the other at λ. If λ is not an integer then the integer part of λ is the mode.

4 If Xi (i=1, 2, …, k) are independent Poisson random variate with parameter λ_i. Then $S_k = \sum_{i=1}^{k} X_i$ has a Poisson distribution with parameter $\sum_{i=1}^{k} \lambda_i$. (This property is called additive or reproductive property).

We have

$$M_{x_i}(t) = e^{\lambda_i(e^t-1)}$$

So

$$M_{s_k}(t) = \prod_{i=1}^{k} M_{x_i}(t) = \prod_{i=1}^{k} e^{\lambda_i(e^t-1)} = e^{\sum \lambda_i(e^t-1)}$$

Which is the mgf of a has a Poisson distribution with parameters $\sum_{i=1}^{k} \lambda_i$.

Hence $S_k = \sum_{i=1}^{k} X_i$ has a Poisson distribution with parameters $(\sum_{i=1}^{k} n_i\ ,\ p)$.

Result

Let X be a Poisson random variable P(λ) and μ_r denote the rth central moment. Then

$$\mu_{r+1} = \lambda \left[\frac{d\mu_r}{dp} + r\mu_{r-1}\right]$$

Proof

We have E(X) = λ and

$\mu_r = E(X-\lambda)^r = \sum_{x=0}^{n} (x-\lambda)^r \frac{\lambda^x}{x!}\ e^{-\lambda}$

So

$$\frac{d\mu_r}{dp} = \sum \quad [x\ \frac{\lambda^{x-1}}{x!}\ e^{-\lambda}(x-\lambda)^r - (x-\lambda)^r \frac{\lambda^x}{x!}\ e^{-\lambda}$$
$$- r\ (x-\lambda)^{r-1} \frac{\lambda^x}{x!}\ e^{-\lambda}$$

$= -r \sum (x-\lambda)^{r-1} \frac{\lambda^x}{x!}\ e^{-\lambda} + \sum (x-\lambda)^{r+1} \frac{\lambda^{x-1}}{x!}\ e^{-\lambda}$

$= -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$

Hence

$$\mu_{r+1} = \lambda \left[\frac{d\mu_r}{dp} + r\mu_{r-1}\right].$$

3. Geometric distribution

Consider a random experiment, which is so defined that it has only two possible outcomes which we call *success* and *failure*. Let p be the probability of a success and q = 1-p be the probability of failure. Let the experiment be repeated till the first success occurs. Also assume that the trails are independent and the probability of success p remains unaltered by trail to trail. Then probability of the number of failure before the first success, say x by multiplication theorem pq^x, where x=0,1,2,…. Such a random variable is said to follow a geometric distribution.

Definition:- A discrete random variable X is said to follow a geometric distribution with parameter p if its probability density function is given by

$$f(x) = pq^x,\ x = 0,1,\ \ldots,\ \ 0 \le p \le 1,\ q = 1-p.$$

Properties

1. **Moments.**

Mean = $\mu_1' = E(X) = \sum_{x=0}^{\infty} x\, pq^x = pq \sum_{x=0}^{\infty} x\, q^{x-1} = pq(1-q)^{-2}$

$$= \frac{pq}{p^2} = \frac{q}{p}.$$

$\mu_2' = E(X^2) = \sum_{x=0}^{\infty} x^2\, pq^x = \sum_{x=0}^{\infty} [x(x-1) + x] p\, q^x$

Now

$$\sum_{x=0}^{\infty} x(x-1) p\, q^x = 2pq^2 \sum_{x=0}^{\infty} \frac{x(x-1)}{2}\, q^{x-2}$$

$$= 2pq^2(1-q)^{-3} = \frac{2q^2}{p^2}$$

$$\mu_2' = \frac{2q^2}{p^2} + \frac{q}{p}.$$

$$V(X) = \mu_2 = E(X^2) - [E(X)]^2 = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$

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

2. Moment generating function

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx}\, p\, q^x = p \sum_{x=0}^{\infty} (qe^t)^x = \frac{p}{1-qe^t}$$

Result

The geometric distribution has the memoryless (forgetfulness) property.

A geometric random variable X has the memoryless property if for all nonnegative integers s and t,

$$P(X \geq s+t | X \geq t) = P(X \geq s)$$

Proof

The probability mass function for a geometric random variable X is

$$f(x) = pq^x, \; x=0,1,2,\dots .$$

Then

$$P(X \geq x) = \sum_{i=x}^{\infty} pq^i = q^x$$

$$P(X \geq s+t | X \geq t) = \frac{P(X \geq s+t \cap x \geq t)}{P(X \geq t)} = \frac{P(X \geq s+t)}{P(X \geq t)}$$

$$= \frac{q^{s+t}}{q^t} = q^s = P(X \geq s)$$

which proves the memoryless property.

$$\text{Mean} = \mu_1' = E(X) = \sum_{x=0}^{\infty} x\, pq^x = pq \sum_{x=0}^{\infty} x\, q^{x-1} = pq(1-q)^{-2}$$

$$= \frac{pq}{p^2} = \frac{q}{p}.$$

$$\mu_2' = E(X^2) = \sum_{x=0}^{\infty} x^2\, pq^x = \sum_{x=0}^{\infty} [x(x-1) + x] p\, q^x$$

Now

$$\sum_{x=0}^{\infty} x(x-1) p\, q^x = 2pq^2 \sum_{x=0}^{\infty} \frac{x(x-1)}{2}\, q^{x-2}$$

$$= 2pq^2(1-q)^{-3} = \frac{2q^2}{p^2}$$

$$\mu_2' = \frac{2q^2}{p^2} + \frac{q}{p}.$$

$$V(X) = \mu_2 = E(X^2) - [E(X)]^2 = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$

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

4. Hyper-Geometric distribution

Consider a lot consisting of *N* items of which *M* of them are defective and the remaining *N* ¡*M* of them are non-defective. A sample of *n* items is drawn randomly without replacement. (That is, an item sampled is not replaced before selecting another item.) Let *X* denote the number of defective items that is observed in the sample. The random variable *X* is referred to as the hypergeometric random variable with parameters *N* and *M*.

Definition: A discrete random variable *X* is said to follow a geometric distribution with parameter *p* if its probability density function is given by

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \text{ x=0, 1, 2 , …, min(n,M)}$$

An alternative form of the pdf is obtained by putting M=Np as

$$f(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, \text{ x=0, 1, 2 , …, min(n,M)}$$

where Np and Nq are integers and p+q=1. In general we take M>n. so that the range of variation is taken as x=0, 1, 2 , …,n.

$$\text{Mean} = \mu_1' = E(X) = \sum_{x=0}^{n} x \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}},$$

Now

$$x\binom{M}{x} = x\frac{M!}{x!(M-x)!} = M\frac{M-1!}{x-1![(M-1)-(x-1)]!} = M\binom{M-1}{x-1}$$

and

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!} = \binom{N}{n}\frac{N-1!}{n-1!\,[(N-1)-(N-1)]!}$$

$$= \frac{N}{n}\binom{N-1}{n-1}.$$

So

Mean = $\mu_1' = \sum_{x=0}^{n} \frac{M\binom{M-1}{x-1}\binom{N-M}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} = \frac{Mn}{N}\sum_{x=0}^{n} \frac{\binom{M-1}{x-1}\binom{N-M}{n-x}}{\binom{N-1}{n-1}} = \frac{Mn}{N}.$

Similarly

$$x(x-1)\binom{M}{x} = M(M-1)\binom{M-2}{x-2}$$

and

$$\binom{N}{n} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}.$$

So

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

$$= \frac{M(M-1)n(n-1)}{N(N-1)}\sum_{x=0}^{n} \frac{\binom{M-2}{x-2}\binom{N-M}{n-x}}{\binom{N-2}{n-2}} = \frac{Mn(M-1)(n-1)}{N(N-1)}$$

μ_2' = E(X^2) = E(X(X-1)) + E(X) = $\frac{Mn(M-1)(n-1)}{N(N-1)} + \frac{Mn}{N}$

hence

V(X) = μ_2 = E(X^2) – [E(X)]^2 = $\frac{Mn(M-1)(n-1)}{N(N-1)} + \frac{Mn}{N} - \left(\frac{Mn}{N}\right)^2$

$$= \frac{Mn(N-n)(N-M)}{N^2(N-1)}.$$