UNIT - II

Hypergeometric distribution

Definition :

A discrete r.v is said to follow the hypergeometric distn with parameters N, M and n if it assumes only non-negative values and its p.m.f is given by

$$p(X=k) = h(k; N, M, n) ;$$

$$\begin{cases} \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} & ; k = 0, 1, 2 \ldots \min(n, M) \quad \text{--- (1)} \\ 0 & \text{otherwise} \end{cases}$$

N → +ve integer

M → +ve integer not exceeding N

n → +ve integer that is atmost N

1. Hypergeometric has 3 parameters M, N, n

2. For hypergeometric distribution the popm should be finite and the sampling ~~is~~ done with replacement so that events are stochastically independent although random.

3. ~~Suppose we draw the sample~~ Suppose the box contains N ball M of them are white N-M are red balls. Suppose we draw the sample

of n balls at random (without replacement) from the box.

The p.m.f given in equation ① gives the probability of getting k white balls out of 'n'

Since k white balls can be drawn from M white balls in $\binom{M}{k}$ and out of the remaining N - M red balls n-k can be chosen in $\binom{N-M}{n-k}$ the total no of fav cases is

$$\binom{M}{k}\binom{N-M}{n-k}$$

Mean of Hypergeometric distribution

$$E(X) = \sum k \, P(X=k)$$

$$= \sum_{k=0}^{n} k \left\{ \binom{M}{k}\binom{N-M}{n-k} \Big/ \binom{N}{n} \right\}$$

$$= \sum k \frac{M!}{k!\,(M-k)!} \binom{N-M}{n-k} \Big/ \binom{N}{n}$$

$$= \frac{M}{\binom{N}{n}} \sum_{k=1} \left\{ \frac{(M-1)!}{(k-1)!\,(M-k)!} \right\} \binom{N-M}{n-k}$$

$$= \frac{M}{\binom{N}{n}} \sum_{k=1}^{n} \binom{M-1}{k-1}\binom{N-M}{n-k} \quad \text{①}$$

substituting $x = k-1$, $m = n-1$, $M-1 = A$

If $k=1 \Rightarrow x=0$

$k=n \Rightarrow x = n-1 = m$

$$= \frac{M}{\binom{N}{n}} \sum_{x=0}^{m} \binom{A}{x} \binom{N-A-1}{m-x}$$

$$= \frac{M}{\binom{N}{n}} \binom{N-1}{m}$$

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

$$= \frac{M}{\frac{N!}{n!\,(N-n)!}} \; \frac{(N-1)!}{(n-1)!\,(N-\cancel{1}-n+\cancel{1})!}$$

$$= \frac{M \times n!\,(N-n)!}{N!} \; \frac{(N-1)!}{(n-1)!\,(N-n)!}$$

$$= \frac{Mn}{N}$$

$V(X) = E(X^2) - [E(X)]^2$

$\cancel{E[X(X-1)]}$ $E(X^2) = \sum [x(x-1)+x]\, p(x)$

$E(X(X-1)) = \sum k(k-1)\, p(x)$

$$= \sum k(k-1) \binom{M}{k} \binom{N-M}{n-k} \Big/ \binom{N}{n}$$

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

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

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

$$= \frac{M(M-1)}{\frac{N!}{n!(N-n)!}} \times \frac{(N-2)!}{(n-2)!\,(N-2-n+2)!}$$

$$= \frac{M(M-1)\,n!\,(N-n)!}{N!} \quad \frac{(N-2)!}{(n-2)!\,(N-n)!}$$

$$= \frac{M(M-1)\,n(n-1)}{N(N-1)}$$

$$E(x^2) = E(x(x-1)) = \frac{M(M-1)\,n(n-1)}{N(N-1)}$$

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

$$= \frac{M(M-1)\,n(n-1)}{N(N-1)} + \frac{nM}{N}$$

$$V(x) = \frac{M(M-1)\,n(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2$$

$$= \frac{M(M-1)\,n(n-1)}{N(N-1)} + \frac{nM}{N} - \frac{n^2M^2}{N^2}$$

$$= \frac{NM(M-1)\,n(n-1) + N(N-1)\,nM - (N-1)(n^2M^2)}{N^2(N-1)}$$

$$\frac{nM[(M-1)\,n(n-1) + N(N-1)] - N\,n^2M^2}{N^2(N-1)}$$

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

Factorial moments of hypergeometric distn

the r^{th} factorial moment is

$$E(X^{r}) = \sum_{k=r}^{n} k^{(r)} \, P(X=k)$$

$$= \sum_{k=r}^{n} k^{(r)} \left\{ \binom{M}{k} \binom{N-M}{n-k} \div \binom{N}{n} \right\}$$

$$k^{(r)} \binom{M}{k} = \frac{k(k-1) \cdots (k-r+1) M!}{k! \, (M-k)!}$$

$$= \frac{M(M-1)(M-2) \cdots (M-r+1) \, (M-r)!}{(k-r)! \, (M-k)!}$$

$$= M^{(r)} \binom{M-r}{k-r}$$

if $j = k-r$

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

as $n^{(r)} \binom{N}{n} = N^{(r)} \binom{N-r}{n-r}$

$$= \frac{M^{r} n^{r}}{N^{r}} \sum_{j=0}^{n-r} h(j;\, N-r,\, M-r,\, n-r)]$$

$$E(X^{r}) = \frac{M^{r} n^{r}}{N^{r}} \cdot 1$$

$$\mu_x = E(X) = \frac{nM}{N} \qquad E(X^{2}) = \frac{M(M-1)n(n-1)}{N(N-1)}$$

Approximation to Binomial distn

hypergeometric distn tends to Binomial distribution as $N \to \infty$ and $\frac{M}{N} \to p$

$$h(k; N, M, n) = \binom{M}{k} \binom{N-M}{n-k} \div \binom{N}{n}$$

$$= \frac{M!}{k!\,(M-k)!} \; \frac{(N-M)!}{(n-k)!\,(N-M-n+k)!} \cdot \frac{n!\,(N-n)!}{N!}$$

$$= \frac{M(M-1)(M-2)\cdots(M-k+1)}{k!}$$

$$\times \frac{(N-M)(N-M-1)\cdots(N-m-n+k+1)}{(n-k)!} \times$$

$$\frac{n!}{N(N-1)(N-2)\cdots(N-n+1)!}$$

$$= \frac{n!}{k!\,(n-k)!} \frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\cdots\left(\frac{M}{N}-\frac{k-1}{n-1}\right)$$

$$\times \frac{\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\cdots\left(1-\frac{M}{N}-\frac{(n+k-1)}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}$$

preceeding to the limit as $N \to \infty$ and putting $\frac{M}{N} \to p$ we get

$$\lim_{N \to \infty} h(k; N, M, n) = \binom{n}{k} \underbrace{p \cdot p \cdot p}_{(k\text{ times})} \underbrace{(1-p)(1-p)\cdots(1-p)}_{(n-k\text{ times})}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

$$= b(k; p, 1-p)$$

Negative Binomial Distribution

The mean and variance is an important characteristic of Poisson distribution.

In Binomial distribution the mean is always greater than the variance.

In some empirical discrete distributions variance is larger than mean. Some of the commonest example are

* Bacterial clustering
* death of insects
* no of insect bites

These above examples leads to negative binomial distribution and the distribution also arises to inverse sampling from a binomial population or as a weighted average of Poisson distribution.

The distn is otherwise known as Pascal distribution.

Consider the Binomial probability situations with some modifications.

Let $f(x; r; p)$ denote that there are x failures preceding the r^{th} success in $x+r$ trials.

The last trial must be a success whose prob is p.

In remaining $(x+r-1)$ trials we must have $(r-1)$ success whose prob is given by binomial prob law

$$\binom{x+r-1}{r-1} p^{r-1} q^{x}$$

Therefore by compound prob theorem $f(x; r, p)$ is given by product of these two probabilities

$$f(x, r, p) = \binom{x+r-1}{r-1} p^{r-1} q^{x} \cdot p$$

$$= \binom{x+r-1}{r-1} p^{r} q^{x} \quad \text{—— (1)}$$

Definition:

A random variable X is said to follow a negative binomial distribution with parameters r and p if its p.m.f is given by

$$P(X=x) = p(x) = \begin{cases} \binom{x+r-1}{r-1} p^{r} q^{x} & ; \ x = 0, 1, 2 \ldots \\ 0 & \text{otherwise} \end{cases}$$

Also $\binom{x+r-1}{r-1} = \binom{x+r-1}{x}$ as $\binom{n}{r} = \binom{n}{n-r}$

$$= \frac{(x+r-1)(x+r-2)\ldots(r+1)r}{x!}$$

$$= \frac{(-1)^{x}(-r)(-r-1)\ldots(-r-x+2)(-r-x+1)}{x!}$$

$$= (-1)^{x}\binom{-r}{x}$$

substituting the above in (1)

$$\therefore \ p(x) = \begin{cases} \binom{-r}{x} p^{x} (-q)^{x} & x = 0, 1, 2 \\ 0 & \text{otherwise} \end{cases} \quad \text{—— (2)}$$

Which is the $(x+1)^{th}$ term in the expansion of $p^{r}(1-q)^{-r}$, a binomial expansion with a negative index. The distn is known as of negative binomial distribution.

Also

$$\sum_{x=0}^{\infty} p(x) = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-q)$$

If $p = \frac{1}{Q}$ $\quad q = \frac{P}{Q}$ so that $Q - P = 1$

$(p+q=1)$ then

$$p(x) = \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x \qquad \text{where } x = 0, 1, 2 \text{ --- } (3)$$

is the general term of negative binomial expansion $(Q-P)^{-r}$

Some important deductions

i) Geometric distn $\quad$ if $r=1$ in (1) then

$p(x) = q^x p \quad ; \; x = 0, 1, 2$

which is the prob fn of geometric distn

ii) Pascal's distn :

The negative binomial distn in (2) when regarded as having 2 parameters p and r is known as Pascal's distribution

$$p(x) = \binom{-r}{x} p^x (-q)^x \qquad x = 0, 1, 2 \ldots$$

iii) Polya's distn

If we take $r = \frac{1}{\beta}$ $\quad p = \frac{1}{1+\beta\mu}$

$q = 1 - p = \frac{\beta\mu}{1+\beta\mu}$ in (3) we get

$$p(x) = \frac{(1+\beta)(1+2\beta)\cdots(1+\beta(x-1))}{x!}\left(\frac{1}{1+\beta\mu}\right)^{1/\beta}\left(\frac{\mu}{1+\beta\mu}\right)^{x}$$

$x = 0, 1, 2 \ldots$

which is known as Polya's distn with paramelar β and μ.

Moment generating functioo

Let X be a negative binomial variable then the m.g.f is defined as

$$M_X(t) = E(e^{tX})$$

$$= \sum e^{tx} p(x)$$

$$= \sum_{x=0}^{\infty} \binom{-r}{x} Q^{-r} \left(\frac{-Pe^t}{Q}\right)^x$$

$$= (Q - Pe^t)^{-r}$$

Mean and variance

Consider the m.g.f of negative Binomial distn

$$M_X(t) = (Q - Pe^t)^{-r}$$

diff w.r.t t

$$\frac{d}{dt} M(t) = -r(-Pe^t)(Q - Pe^t)^{-r-1}$$

at $t = 0$

$$\left.\frac{d}{dt} M(t)\right|_{t=0} = \mu_1' = \left. -r(-Pe^t)(Q - Pe^t)^{-r-1} \right|_{t=0}$$

$$= rPe^0 (Q - Pe^0)^{-r-1}$$

$$= rP(Q - P)^{-r-1}$$

$$\mu_1' = rP$$

as $Q - P = 1$

$$\frac{d}{dt} M_x(t) = r p e^t (Q - Pe^t)^{-r-1}$$

$$\frac{d^2 M_x(t)}{dt^2} = rpe^t (Q - Pe^t)^{-r-1} + (-r-1) rpe^t (Q - Pe^t)^{-r-2} (-Pe^t)$$

sub $t = 0$

$$\frac{d^2 M_x(t)}{dt^2} = \mu_2' = rpe^0 (Q - Pe^0)^{-r-1} + (-r-1) rpe^0 (Q - Pe^0)^{-r-2} (-Pe^0)$$

$$= rp(Q-P)^{-r-1} + (-r-1) rP (Q-P)^{-r-2} (-P)$$

$$\mu_2' = rp + r(r+1)P^2$$

Variance

$$\mu_2 = \mu_2' - (\mu_1')^2$$
$$= rp + r(r+1)p^2 - r^2p^2$$
$$= rP + \cancel{r^2}p^2 + rp^2 - \cancel{r^2}p^2$$
$$= rP(1-P)$$
$$= rPQ$$

As $Q > 1$, $rp < rPQ$

$\Rightarrow$ mean < variance which is the distinguishing feature of negative Binomial distr.

cumulants

$$K_x(t) = \log M_x(t)$$
$$= -r \log (Q - Pe^t)$$
$$= -r \log \left[Q - P \left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots \right)\right]$$

$K_x(t) = -r\log\left[1 - P\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots\right)\right]$

replacing n with $-r$ and p with $-p$ we get

mean $= K_1 = rP$

$\mu_2 = rP(1+P) = rPQ$

$\mu_3 = k_3 = rP(1+3P+2P^2)$
$= rP(1+P)(1+2P)$
$= rPQ(Q+P)$

$K_4 = rP(1+P)(1+6P+6P^2)$
$= rPQ(1+6PQ)$

$\mu_4 = K_4 + 3_K{}^2$
$= rPQ[1+3PQ(r+2)]$

since $Q = 1/p$; $P = qQ = q/p$

Mean $= \frac{rq}{p}$

Variance $\mu_2 = \frac{rq}{p^2}$

$\mu_3 = \frac{rq(1+q)}{p^3}$

$\mu_4 = \frac{rq(p^2+3q(r+2))}{p^4}$

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(1+q)^2}{rq}$ $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{p^2+3q(r+2)}{rq}$

$\gamma_1 = \sqrt{\beta_1} = 1+q/\sqrt{rq}$ $\gamma_2 = \beta_2 - 3$
$= (p^2+6q)/rq$

Geometric distribution

Consider a series of independent trials or repetition and in each trial prob of success remains the same. Then the prob that there are x failure preceding the first success is given by $q^x \cdot p$, $q = 1-p$

Defn:

A random variable X is said to have a geometric distribution if it assume only non-ve values and its p.m.f is given by

$$P(X=x) = \begin{cases} q^x p & x = 0,1,2\ldots \quad 0 < p \le 1 \\ 0 & \text{otherwise} \quad q = 1-p \end{cases}$$

Remark:

* The various probabilities for $x = 0, 1, 2 \cdots$ are the terms of geometric progression series hence the name geometric distn

* The probabilities are permissible since

$$\sum_{x=0}^{\infty} P(X=x) = \sum_{x=0}^{\infty} q^x p$$

as $(1-x)^{-1} = (1+x+x^2+\cdots)$

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

$$= \frac{p}{1-q}$$

$$= p(p)^{-1}$$

$$= 1$$

Lack of memory:

Geometric distribution is said to possess lack of memory in a certain sense.

Suppose an event E can occur at one of the time $t = 0, 1, 2 \ldots$

The occurrence (waiting) time X has a geometric distn with parameter p.

Thus

$$P(X = t) = q^t . p \quad ; \quad t = 0, 1, 2$$

Suppose the event has not occured before k ie $X > k$.

Let $y = X - k$ where y is the amount of additional time needed for E to occur

$$P(y = t / X \geq k) = P(X = t) = pq^t \quad \text{——} \quad (1)$$

which implies that the additional time to wait

Since the distribution does not depend upon k, it in a sense, lacks memory' of how much we shifted the time origin.

If 'B' were waiting for the event E and relieved by c immediatly before time k, then the waiting time distn of C is the same as that of B

Proof :

To prove $P(y = t / X \geq k) = P(X = t) = pq^t$

$$P(X \geq r) = \sum_{s=r}^{\infty} pq^s$$

$$= p(q^r + q^{r+1} + q^{r+2} + \ldots)$$

$$= pq^r(1 + q + q^2 + \ldots)$$

$$= pq^r(1-q)^{-1}$$

$$= \frac{pq^r}{1-q} = \frac{pq^r}{p} = q^r \quad \text{——} \quad (2)$$

$$P(y \geq t / x \geq k) = \frac{P(y \geq t \cap x \geq k)}{P(x \geq k)}$$

$$= \frac{P(x-k \geq t \cap x \geq k)}{P(x \geq k)} \qquad \because y = x-k$$

$$= \frac{P(x \geq k+t)}{P(x \geq k)}$$

$$= \frac{q^{t+k}}{q^k} = q^t \quad \text{— ②}$$

$$\therefore P(y = t / x \geq k) = P(y \geq t / x \geq k) - P(y \geq t+1 / x \geq k)$$

$$= q^t - q^{t+1}$$

$$= q^t (1-q)$$

$$= p q^t$$

$$= P(x = t) \quad \text{from ②.}$$

Moments of geometric distn

$$E(x) = \mu_1' = \sum_{x=0}^{\infty} x\, p(x)$$

$$= \sum x q^x p$$

$$= pq \sum x q^{x-1}$$

$$= pq (1-q)^{-2}$$

$$= \frac{q}{p} = qp^{-1}$$

as $\sum q^{x-1} = 1+2q+3q^2+4q^3 = (1-q)^{-2}$

$$E(x^2) = \sum x^2 p(x)$$

$$= \sum [x(x-1)+x]\, p(x)$$

$$= \sum x(x-1) pq^x + \sum x pq^x$$

$$= 2pq^2 \sum_{x=2}^{\infty} \frac{x(x-1)}{2\times 1} q^{x-2} + \dots + qp^{-1}$$

$$= 2pq^2 \left[1 + \frac{3(3-1)}{2} \times q^{3-2} + \frac{4(4-1)}{2} q^{4-2} + \dots \right] + qp^{-1}$$

$$= 2pq^2 [1 + 3q + 6q^2 + \dots] + qp^{-1}$$

$$= 2pq^2 (1-q)^{-3} + qp^{-1}$$

$$= \frac{2pq^2}{p^3} + \frac{q}{p} = \frac{2q^2}{p^2} + \frac{q}{p}$$

$$V(X) = E(X^2) - [E(X)]^2$$

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

$$= \frac{2q^2 + qp - q^2}{p^2}$$

$$= \frac{q^2 + qp}{p^2}$$

$$= \frac{q(q+p)}{p^3}$$

$$V(X) = qp^{-2}$$

Moment generating function

Let X be geometric random variable and its m.g.f is defined by

$$M_x(t) = E[e^{tx}]$$

$$= \sum e^{tx} p(x)$$

$$= \sum e^{tx} q^x p$$

$= \sum e^{tx} q^x p$

$= p \sum_{x=0}^{\infty} (e^t q)^x$

$= p [1 + qe^t + (qe^t)^2 + \dots]$

$= p(1-qe^t)^{-1}$

$= \dfrac{p}{1-qe^t}$

From this function we can find mean and variance by diff w.r.t 't'

$\mu_1' = \dfrac{d}{dt} M_x(t)$

$= \dfrac{d}{dt} p(1-qe^t)^{-1} \Big|_{t=0}$

$= -p(1-qe^t)^{-2}(-qe^t) \Big|_{t=0}$

$= pqe^t(1-qe^t)^{-2} \Big|_{t=0}$

$= pq(1-q)^{-2}$

$= pq(p)^{-2}$

$= \dfrac{q}{p}$

$\mu_2 = \dfrac{d}{dt} pqe^t(1-qe^t)^{-2}$

$pq\left[e^t(-2)(1-qe^t)^{-3}(-qe^t) + (1-qe^t)^{-2} e^t\right]_{t=0}$

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

$= pq(2qp^{-3} + p^{-2})$

$$= pq(2qp^{-3} + p^{-2})$$

$$= 2q^2p^{-2} + qp^{-1}$$

$$= \frac{2q^2}{p^2} + \frac{q}{p}$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$= \frac{2q^2}{p^2} + \frac{q}{p} - (qp^{-1})^2$$

$$= \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= \frac{2q^2 + pq - q^2}{q^2}$$

$$= q^2 + qp / q^2$$

$$= q(q+p) / p^2$$

$$\mu_2 = \frac{q}{p^2}$$

$\therefore$ Mean $= \dfrac{q}{p}$

Variance $= \dfrac{q}{p^2}$