Discrete Prob Distributions

1. Discrete Uniform Distribution:

Def^n: A discrete r.v. X is said to have a discrete uniform prob distribution over the range $[1, n]$, if its p.m.f. is expressed as follows:

$$P(X = x) = p(x) = \frac{1}{n}, \quad x = 1, 2, \ldots n$$
$$= 0, \quad \text{otherwise}$$

Here n is known as the parameter of the distribution and lies in the set of all positive integers. This distribution is also called a discrete rectangular distribution.

Situation when this dist^n is applicable: If under the given experimental conditions, the different values of the random variable become equally likely. Thus for a die experiment, and for an experiment with a deck of cards such dist^n is appropriate.

Moments:

$$\mu_1' = E(X) = \sum_{x=1}^{n} x\, p(x) = \frac{1}{n}\sum_{x=1}^{n} x = \frac{1}{n}\frac{n(n+1)}{2} = \frac{(n+1)}{2}$$

$$\mu_2' = E(X^2) = \sum_{x=1}^{n} x^2 p(x) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{1}{n}\frac{n(n+1)(2n+1)}{6}$$
$$= \frac{(n+1)(2n+1)}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{(n+1)(n-1)}{12}$$

The MGF of X is $M_X(t) = E e^{tX} = \sum_{x=1}^{n} e^{tx} p(x)$

$$= \frac{1}{n}\sum_{x=1}^{n} e^{tx} = \frac{e^t}{n}\left[\frac{(1-e^{nt})}{(1-e^t)}\right]$$

Negative Binomial Distribution (NBD):

In $B(n, p)$, Mean > Variance

In $P(\lambda)$, Mean = Variance

In NBD Mean < Variance

Application of NBD: Ex. Bacterial clustering or contagion e.g. deaths of insects, number of insects bites leads to the NBD.

Suppose we have a succession (series) of n Bernoulli trials. We assume that i) the trials are independent, (ii) the probability of success 'p' in a trial remains constant from trial to trial.

Let $f(x; r, p)$ denote the probability that there are x failures preceeding the rth success in $x+r$ trials. Now, the last trial must be a success, whose probability is p. In the remaining $(x+r-1)$ trials we must have $(r-1)$ success whose prob is given by the binomial prob law by the expression : $\binom{x+r-1}{r-1} p^{r-1} q^{x}$.

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

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

$x = 0, 1, 2, \ldots$

Definition : A random variable X is said to follow a negative binomial distribution with parameters r and p if its pmf is given by

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

Also $$\binom{x+r-1}{r-1} = \binom{x+r-1}{x} \qquad \left[\binom{n}{r} = \binom{n}{n-r} \right]$$

$$= \frac{\lfloor x+r-1}{\lfloor x \lfloor r-1} = \frac{(x+r-1)(x+r-2)\cdots(r+1)r\cancel{\lfloor r-1}}{\lfloor x \cancel{\lfloor r-1}}$$

$$= \frac{(-1)^{x}(-r)(-r-1)\cdots(-r-x+2)(-r-x+1)}{\lfloor x} = (-1)^{x}\binom{-r}{x}$$

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

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

Also $$\sum_{x=0}^{\infty} p(x) = p^{r} \sum_{x=0}^{\infty} \binom{-r}{x}(-q)^{x} = p^{x}(1-q)^{-r} = 1.$$

Hence $p(x)$ is pmf.

If $p = \frac{1}{Q}$ and $q = \frac{P}{Q}$, so that $Q - P = 1$, ($\because p + q = 1$), then

$$p(x) = \begin{cases} \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x ; & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

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

Remark: (1) If we take $r = 1$ in ⊛, we have

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

which is the prob function of geometric dist^n. Hence NBD may be regarded as the generalization of geometric dist^n.

Moment Generating Function of Negative Binomial Distribution.

$$M_X(t) = E\, e^{tx} = \sum_{x=0}^{\infty} e^{tx} p(x) = \sum_{x=0}^{\infty} \binom{-r}{x} Q^{-r} \left(-\frac{Pe^t}{Q}\right)^x$$

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

$$\text{Mean } \mu_1' = \frac{d}{dt} M_X(t) \Big|_{t=0} = \left[-r(-Pe^t)(Q-Pe^t)^{-r-1}\right]_{t=0} = rP$$

$$\text{Mean} = rP$$

$$\mu_2' = \frac{d^2 M_X(t)}{dt^2}\Big|_{t=0} = \left[rPe^t(Q-Pe^t)^{-r-1} + (-r-1)rPe^t(Q-Pe^t)^{-r-2}(-Pe^t)\right]_{t=0}$$

$$= rP + r(r+1)P^2$$

Variance $\therefore \mu_2 = \mu_2' - \mu_1'^2 = rP + r(r+1)P^2 - r^2P^2 = rPQ$

As $Q > 1$, $rP < rPQ$ i.e. Mean < Variance which is distinguish feature of the NBD.

Ex: An item is produced in large numbers. The machine is known to produce 5% defective. A quality control inspector is examining the items by taking them at random. What is the probability that at least 4 items are to be examined in order to get 2 defectives?

Sol^n: If 2 defectives are to be obtained then it can happen in 2 or more trials. The prob. of success is 0.05 for every trial. It is negative binomial ~~distb~~ situation with $r=2$, $p=0.05$ and x replaced by $(x-2)$ and the required prob is given by

$$P(X=4)+P(X=5)+\cdots = \sum_{x=4}^{\infty}\binom{x-1}{2-1}(0.05)^2(0.95)^{x-2}$$

$$= 1-\sum_{2}^{3}\binom{x-1}{2-1}(0.05)^2(0.95)^{x-2} = 0.9928.$$

$x+r \geq 4$

$x+r = U$

$x = U-r = U-2$

$$P(U\geq 4) = \sum_{U=4}^{\infty}\binom{U-1}{2-1}(0.05)^2(0.95)^U$$

$=$

Geometric Distribution:

Suppose we have a series of independent bernoulli trials or repetitions and in each trial the prob of success 'p' remains the same. Then the probability that there are x failures preceding the first success is given by $q^x p$, $q=1-p$.

Def^n: A random variable X is said to have a geometric distribution if it assumes only non-negative values and its prob mass function is given by

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

$$0<p\leq 1;\ q=1-p.$$

Remark: (i) The various probabilities for $x=0,1,2,\dots$ are the terms of geometric series, hence the name geometric dist^n.

(ii) $\sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} q^x p = p(1+q+q^2+\cdots) = \frac{p}{1-q} = \frac{p}{p} = 1$, hence its pmf.

Lack of Memory: The geometric distribution is said to lack memory in a certain sense. Suppose an event E can occur at one of the times $t = 0, 1, 2, \ldots$ and the occurrence (waiting) time X has a geometric distribution with parameter p

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

Suppose we know that the event E has not occurred before k, i.e, $X \geq k$. Let $Y = X - k$. Thus, Y is the amount of additional time needed for E to occur. We can show that

$$P(Y = t \mid X \geq k) = P(X = t) = pq^t$$

which implies that the additional time to wait has the same distn as initial time to wait.

i.e Since the distribution does not depend k, it, in a sense, 'lack memory' of how much we shifted the time origin. If 'B' were waiting for the event E and is relieved by 'C' immediately before time k, then the waiting time distribution of 'C' is the same as that of 'B'.

Proof: $$P(X \geq r) = \sum_{s=r}^{\infty} pq^s = p(q^r + q^{r+1} + q^{r+2} + \cdots) = \frac{pq^r}{(1-q)} = q^r \quad \text{--- } (*)$$

$$P(Y \geq t \mid X \geq k) = \frac{P(Y \geq t \cap X \geq k)}{P(X \geq k)} = \frac{P(X - k \geq t \cap X \geq k)}{P(X \geq k)}$$

$$= \frac{P(X \geq k + t)}{P(X \geq k)} = \frac{q^{t+k}}{q^k} = q^t \quad \text{--- } (**) \quad (\because Y = X - k)$$

(From $(*)$)

$$\therefore P(Y = t \mid X \geq k) = P(Y \geq t \mid X \geq k) - P(Y \geq t+1 \mid X \geq k)$$

$$= q^t - q^{t+1} = q^t(1-q) = pq^t = P(X = t)$$

(From $(**)$)

Moment Generating Function of Geometric Distribution

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

$$= \frac{p}{1-qe^t}.$$

$$\text{Mean } \mu_1' = \frac{d}{dt} M_X(t)\Big|_{t=0} = \left[\frac{d}{dt} p(1-qe^t)^{-1}\right]_{t=0} = pq(1-q)^{-2} = \frac{q}{p}$$

$$\mu_2' = \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = \frac{q}{p} + \frac{2q^2}{p^2} \quad \text{(after simplification)}$$

$$\therefore \text{ Variance } \mu_2 = \mu_2' - \mu_1'^2 = \frac{q}{p} + \frac{2q^2}{p^2} - \frac{q^2}{p^2} = \frac{q}{p} + \frac{q^2}{p^2} = \frac{pq+q^2}{p^2}$$

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

Hence the mean and variance of geometric distribution are $\frac{q}{p}$ and $\frac{q}{p^2}$ respectively.

Remark: $\text{Variance} = \frac{q}{p^2} = \frac{q}{p} \cdot \frac{1}{p} = \frac{\text{Mean}}{p} > \text{Mean}$

Hence, in geometric distⁿ Variance > Mean.

Ex. Let X_1, X_2 be independent r.vs each having geometric distribution $q^k p$: $k = 0, 1, 2 \ldots$ Show that the conditional distribution of X_1 given $(X_1 + X_2)$ is Uniform.

Solⁿ. $$P(X_1 = r \mid X_1 + X_2 = n) = \frac{P(X_1 = r \cap X_1 + X_2 = n)}{P(X_1 + X_2 = n)}$$

$$= \frac{P(X_1 = r \cap X_2 = n-r)}{P(X_1 + X_2 = n)} = \frac{P(X_1 = r \cap X_2 = n-r)}{\sum_{s=0}^{n} P(X_1 = s \cap X_2 = n-s)}$$

$$= \frac{P(X_1 = r)\, P(X_2 = n-r)}{\sum_{s=0}^{n}\left[P(X_1 = s)\, P(X_2 = n-s)\right]}$$ Since X_1 and X_2 are independent.

$$\therefore P[X_1 = r \mid (X_1 + X_2 = n)] = \frac{pq^r \cdot pq^{n-r}}{\sum_{s=0}^{n}\left[pq^s \cdot pq^{n-s}\right]} = \frac{p^2 q^n}{\sum_{s=0}^{n} p^2 q^n} = \frac{p^2 q^n}{(n+1)p^2 q^n} = \frac{1}{n+1};\ r = 0, 1, \ldots n.$$

Hence the result

Ex Suppose X is a non-negative integral random variable. Show that the distribution of X is geometric if it 'lacks memory' i.e. if for each $K \geq 0$ and $Y = X-K$ one has $P(Y=t \mid X \geq K) = P(X=t)$, for $t \geq 0$.

Sol^n. Let us suppose $P(X=r) = p_r$; $r = 0, 1, 2 \ldots$ —— (i)

Define $q_K = P(X \geq K) = p_K + p_{K+1} + p_{K+2} \ldots$ —— (ii) (*)

We are given $P(Y=t \mid X \geq K) = P(X=t) = p_t$ —— (iii)

$$P(Y=t \mid X \geq K) = \frac{P(Y=t \cap X \geq K)}{P(X \geq K)} = \frac{P(X-K=t \cap X \geq K)}{P(X \geq K)}$$

$$= \frac{P(X=t+K)}{P(X \geq K)} = \frac{p_{K+t}}{q_K}$$

$\therefore$ $p_t = \dfrac{p_{K+t}}{q_K}$, for every $t \geq 0$ and all $K \geq 0$ —— From (iii)

In particular, taking $K=1$, we get

$p_{t+1} = q_1 p_t = (p_1 + p_2 + \cdots) p_t = (1-p_0) p_t$ From (*)

$\Rightarrow p_t = (1-p_0) p_{t-1} = (1-p_0)^2 p_{t-2} = \cdots = (1-p_0)^t p_0$

Hence $p_t = P(X=t) = p_0 (1-p_0)^t$; $t = 0, 1, 2, \ldots$

$\Rightarrow$ X has a geometric distribution.

Hyper. Geometric Distribution:

Consider an urn with N balls, M of which are white and $N-M$ are red. Suppose that we draw a sample of n balls at random (without replacement) from the urn. Then the prob of getting K white balls out of n, $(K < n)$ is $\dfrac{\binom{M}{K}\binom{N-M}{n-K}}{\binom{N}{n}}$

[Since K white balls can be drawn from 'M' white balls in $\binom{M}{K}$ ways and out of the remaining $N-M$ red balls, $(n-K)$ balls can be chosen in $\binom{N-M}{n-K}$ ways, the total number of favourable cases is $\binom{M}{K}\binom{N-M}{n-K}$]

Defn: A discrete r.v X is said to follow the hypergeometric distribution with parameters N, M and n if it assumes only non-negative values and its prob mass function is given by:

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

Where N is a positive integer, M is a positive integer not exceeding N and n is a positive integer that is at most N.

As it can be shown that $\sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} \Big/ \binom{N}{n} = 1$.

Mean $$E(X) = \sum_{k=0}^{n} k P(X=k) = \sum_{k=0}^{n} k \left\{ \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \right\}$$

$$= \frac{nM}{N}$$

$$E(X^2) = \sum_{k=0}^{n} k^2 P(X=k) = \sum_{k=0}^{n} \{k(k-1)+k\} P(X=k)$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} \quad \text{(on simplification)}$$

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

(on simplification)