BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Distribution Theory

Course Code: 23ST02CC

Unit-III

Discrete distribution

Dr. T. Jai Sankar

Associate Professor and Head

Department of Statistics

Ms. J. Jenitta Edal Queen

Guest Faculty

Department of Statistics

Unit III

Geometric Distribution

Definition

A random variable X is said to have the geometric distribution if its assumes only non-negative values, and its probability mass function is given by

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

Remark

Since the various probability for $x = 0,1,2,$....are the various terms of geometric progression. Hence, that named as geometric distribution

Applications of Geometric Distribution

- Geometric distribution has a wide range of applications in real life. The geometric distribution can be applied on an intuitive level in our daily lives, regularly. We all have been there when we are trying to reach our destination and all the traffic lights are red. The geometric distribution can help us find the probability of getting to the green light before reaching the destination.
- Geometric distribution can help find the probability of how many red lights are needed to get to a green light before reaching a destination
- In real-world applications, the geometric distribution is used to examine the possibility of success given a limited number of trials, where the probability of success is small. Some examples are identifying an infected person who caused an epidemic in a ward containing 100 patients or estimating the mean number of coin flips required to obtain heads for the first time.
- The geometric distribution is an important statistical concept that is applied in various fields such as baseball, time management, and cost-benefit analysis.
- The geometric distribution is used to model things like the probability of flipping a coin until you get heads or the probability of throwing a die until you get a six. It can also be used to model the probability of birth control working per month.

Properties of Geometric Distribution

The geometric distribution is basically a discrete probability distribution and this distribution help to determine the number of trials before getting the first success. The mean, mode and variance are the three values that are generally used for geometric distribution whereas the median is not generally determined.

- A geometric distribution is a discrete probability distribution that shows the probability of a certain number of events with a certain probability before another event occurs.
- The mean of geometric distribution shows the expected value of the distribution
- The mode is the highest occurrence for a given set of data.
- Generally, mean, mode and variance are used for geometric distribution whereas the median is not computed.

- The expected value of the geometric distribution is the number of trials needed to get success.
- The geometric distribution also known as the negative binomial distribution is a discrete probability distribution. It's most commonly associated with the number of trials required to get the first success in a sequence of Bernoulli trials.

Clearly assignment of probability is the above probability mass function is permissible, since

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

$$= p \sum_{x=1}^{n} q^x$$

$$= p[1 + p + p^2 + \cdots$$

$$= p[1 - q]^{-1}$$

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

$$= \frac{p}{p}$$

$$\sum_{x=0}^{\infty} p(x) = 1$$

Moments of Geometric Distribution

Mean

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

$$= \sum_{x=0}^{\infty} x\, q^x p$$

$$= p \sum_{x=0}^{\infty} x\, q^x$$

$$= p \sum_{x=1}^{\infty} x\, q^{x-1} q$$

$$= pq \sum_{x=1}^{\infty} x\, q^{x-1}$$

$= pq[1 + 2q + 3q^2 + \cdots]$

$= pq[1 - q]^{-2}$

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

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

$\mu_1^{'} = \frac{q}{p}$

Variance $\mu_2 = E(X^2) - [E(X)]^2$

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

$= \sum_{x=2}^{\infty} x(x-1)p(x) + \sum_{x=1}^{\infty} x\, p(x)$

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

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

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

$= pq^2[2 + 6q + 12q^2 + \cdots] + \frac{q}{p}]$

$= 2pq^2[1 + 3q + 6q^2 + \cdots] + \frac{q}{p}]$

$= 2pq^2[1 - q]^{-3} + \frac{q}{p}$

$= \frac{2pq^2}{[1-q]^3} + \frac{q}{p}$

$$E(X^2) = \frac{2q^2}{p^2} + \frac{q}{p}$$

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

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

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

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

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

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

Standard Deviation of Geometric Distribution

As we know, the standard deviation is defined as the square root of the variance. The standard deviation provides us with the deviation of the distribution with respect to the mean. The standard deviation of a geometric distribution can be calculated using the formula which is as follows:

$$\mathbf{S.D.} = \sqrt{\boldsymbol{var(X)}}$$

$$\mathbf{S.D.} = \sqrt{\frac{\boldsymbol{(1-p)}}{\boldsymbol{p^2}}} = \frac{\sqrt{\boldsymbol{1-p}}}{\boldsymbol{p}}$$

Moment Generating Function of Geometric Distribution

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

$$= \sum_{x=0}^{\infty} e^{tx}\, p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx}\ q^x p$$

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

$$= p[1 + (e^t q)^1 + (e^t q)^2 + \cdots]$$

$$= p[1 - e^t q]^{-1}$$

$$M_X(t) = \frac{p}{1 - e^t q}$$

$$\mu_1' = \left[\frac{d}{dt} M_X(t)\right]_{t=0}$$

$$= \left[\frac{d}{dt} p[1 - e^t q]^{-1}\right]_{t=0}$$

$$= [p(1 - qe^t)^{-2}(-qe^t)]_{t=0}$$

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

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

$$\mu_1^{'} = \frac{q}{p}$$

$$\mu_2^{'} = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0}$$

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

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

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

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

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

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

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

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

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

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

Lack of Memory

The geometric distribution is said to lack of memory in a certain sense suppose an event E can occur at one of the times t=0,1,2,….. and the occurrence (waiting) time x has a geometric distribution with parameter p.

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

Suppose w.r.t the event E has not occurred before k,(i.e,) X ≥ k. Let Y= X-k. Thus Y is the amount of additional time needed for E to occur . We can S.T

$$p(Y = t\ / X \geq k) = p(X = t) = pq^t \to (1)$$

Which implies that the additional time to wait has the same distribution as initial time to wait.

Since the distribution does not depend upon 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

We have $p(X \geq r) = \sum_{x=r}^{\infty} pq^x$

$= p[q^r + q^{r+1} + q^{r+2} + \cdots]$

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

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

$= \dfrac{pq^r}{1 - q}$

$= \dfrac{pq^r}{p}$

$$p(X \geq r) = q^r \to (2)$$

$p(Y \geq t\ / X \geq k) = \dfrac{p(Y \geq t \cap X \geq k)}{p(X \geq k)}$

Put Y=X-k

$= \dfrac{p(X - K \geq t \cap X \geq k)}{p(X \geq k)}$

$= \dfrac{p(X \geq k + t \cap X \geq k)}{p(X \geq k)}$

$(X \geq k) = k, k + 1, k + 2, \ldots\ldots k + t, \ldots\ldots$

$X \geq k + t = k + t, \ldots\ldots$

$= \dfrac{p(X \geq k + t)}{p(X \geq k)}$

$= \dfrac{q^{k+t}}{q^k}$

$$= \frac{q^k q^t}{q^k}$$

$$p(Y \geq t\ /X \geq k) = q^t \rightarrow (3)$$

$$p(Y = t\ /X \geq k) = p(Y \geq t\ /X \geq k) - p(Y \geq t+1\ /X \geq k)$$

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

$$= q^t - q^t q$$

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

$$= pq^t$$

$$= p(X = t)$$

$$p(Y = t\ /X \geq k) = p(X = t) = pq^t$$

Hence proved

Negative Binomial Distribution

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

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

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

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

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

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

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

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

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

$$p = \frac{1}{Q}, q = \frac{P}{Q}. Q - P = 1$$

$$p(x) = \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x\ ; x = 0,1,2..$$

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

Properties of Negative Binomial Distribution

A negative binomial distribution is a distribution that has the following properties.

- The negative binomial distribution has a total of n number of trials.
- Each trial has two outcomes, and one of them is referred to as success and the other as a failure.
- The probability of success or failure is the same across each of these trials.
- The probability of success is denoted by p, and the probability of failure is defined as q, and each of these is the same in every trial.
- The sum of the probability of success and failure is equal to 1. P + q = 1.
- Each of these trials is independent. The outcome of one trial does not affect the outcome of other trials.
- The experiment is continued until r success is obtained, and r is defined in advance.
- The experiment consists of x + r repeated trials, where r is the required number of successes.

Examples of Negative Binomial Distribution

The following quick examples help in a better understanding of the concept of the negative binomial distribution.

- If we flip a coin a fixed number of times and count the number of times the coin turns out heads is a binomial distribution. If we continue flipping the coin until it has turned a particular number of heads say the third head-on flipping 5 times, then this is a case of the negative binomial distribution.
- For a situation involving three glasses to be hit with 7 balls, the probability of hitting the third glass successfully with the seventh ball can be obtained with the help of negative binomial distribution.
- In a class, if there is a rumor that there is a math test, and the fifth is the second person to believe the rumor, then the probability of this fifth person to be the second person to believe the rumor can be computed using the negative binomial distribution.

Moment Generating Function of Negative Binomial Distribution

If X ~ N.B.D and if P.M.F is

$$p(x) = \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^{x} \ ; x = 0,1,2..$$

Now, the m.g.f of x is

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

$$= \sum_{x=0}^{\infty} e^{tx}\, p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x$$

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

$$= \left[\binom{-r}{0} Q^{-r} \left(-\frac{P}{Q} e^t\right)^0 + \binom{-r}{1} Q^{-r} \left(-\frac{P}{Q} e^t\right)^1 + \binom{-r}{2} Q^{-r} \left(-\frac{P}{Q} e^t\right)^2 + \cdots \ldots\right]$$

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

Moments

$$\mu_1^{'} = \left[\frac{d}{dt} M_X(t)\right]_{t=0}$$

$$= \left[\frac{d}{dt} (Q - Pe^t)^{-r}\right]_{t=0}$$

$$= [-r(Q - Pe^t)^{-r-1}(-Pe^t)]_{t=0}$$

$$= [rPe^t(Q - Pe^t)^{-r-1}]_{t=0}$$

$$= rPe^0(Q - Pe^0)^{-r-1} \;\; [e^0 = 1]$$

$$\mu_1^{'} = rP$$

$$\mu_2^{'} = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0}$$

$$= \left[\frac{d^2}{dt^2} (Q - Pe^t)^{-r}\right]_{t=0}$$

$$= \frac{d}{dt}\left[\frac{d}{dt} (Q - Pe^t)^{-r}\right]_{t=0}$$

$$= \frac{d}{dt} [-r(Q - Pe^t)^{-r-1}(-Pe^t)]_{t=0}$$

$$= \frac{d}{dt} [rPe^t(Q - Pe^t)^{-r-1}]_{t=0}$$

$$= [rPe^t(-r - 1)(Q - Pe^t)^{-r-2}(-Pe^t) + rPe^t(Q - Pe^t)^{-r-1}]_{t=0}$$

$$= rP(-r - 1)(-P) + rP$$

$$= rP + rP(rP + P)$$

$$\mu_2^{'} = rP + rP^2(r + 1)$$

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

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

$= rP + r^2P^2 + rP^2 - r^2P^2$

$= rP + rP^2$

$= rP(1+P)$

$$\mu_2 = rPQ$$

Cumulates of Negative Binomial Distribution

If X ~ N.B.D and if P.M.F is

$$p(x) = \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x \ ; x = 0,1,2..$$

We know that

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

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

Mean

$\mu_1' = k_1$ is co efficient of t is $k_x(t)$

$$k_x(t) = -r(-P) = rP$$

$\mu_2 = k_2$ is co efficient of $\frac{t^2}{2!}$ is $k_x(t)$

$$k_x(t) = rP - rP^2$$

$$= rP(1-P)$$

$$k_2 = rPQ$$

$\mu_3 = k_3$ is co efficient of $\frac{t^3}{3!}$ is $k_x(t)$

$$k_x(t) = rP + 2rP^3 + 3rP^2$$

$= rP(1 + 2P^2 + 3P)$

$= rP(1 - P)(1 + 2P)$

$= rPQ(1 + 2P)$

$= rPQ(Q - P + 2P)$

$k_3 = rPQ(Q + P)$

k_4 is co efficient of $\frac{t^4}{4!}$ is $k_x(t)$

$$k_x(t) = rP\left(\frac{t^4}{4!} + \frac{rP^2}{2!}\frac{t^4}{4} + \frac{2rP^2}{2}\frac{t^4}{6} + \frac{3rP^3}{3}\frac{t^4}{2} + \frac{rP^4}{4}\frac{t^4}{1!}\right)$$

$$= \left(rP\frac{t^4}{4!} + \frac{3rP^2t^4}{4!} + \frac{4rP^2t^4}{4!} + \frac{12rP^3t^4}{4!} + \frac{6rP^4t^4}{4!}\right)$$

$= rP + 7rP^2 + 12rP^3 + 6rP^4$

$= rP(1 + P)(1 + 6P + 6P^2)$

$= rPQ[(1 + P)(1 + 6P)]$

$$k_4 = rPQ[1 + 6PQ]$$

$$\mu_4 = k_4 + 2{k_2}^2$$

$= rPQ[1 + 6PQ] + 3\,\{rPQ\}^2$

$= rPQ[1 + 6PQ + 3rPQ]$

$$\mu_4 = rPQ[1 + 3PQ(r + 2)]$$

Since

$$Q = \frac{1}{p}, P = qQ = \frac{q}{p}$$

We have

$$\mu_1 = mean = rP = \frac{rq}{p}$$

$$\mu_2 = variance = rPQ = \frac{rq}{p}\frac{1}{p} = \frac{rq}{P^2}$$

$$\mu_3 = rPQ[Q + P]$$

$$= \frac{rq}{p}\left(\frac{1}{p}\right)\left[\frac{1}{p} + \frac{q}{p}\right]$$

$$= \frac{rq}{P^2}\left[\frac{1 + q}{p}\right]$$

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

$$\mu_4 = rPQ[1 + 3PQ(r + 2)]$$

$$= \frac{rq}{p}\left(\frac{1}{p}\right)\left[1 + \frac{1}{p}\frac{3q}{p}(r+2)\right]$$

$$= \frac{rq}{P^2}\left[1 + \frac{3q}{P^2}(r+2)\right]$$

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

$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}$$

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

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

$$\beta_1 = \frac{(1+q)^2}{rq}$$

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2}$$

$$= \frac{\frac{rq[P^2 + 3q(r+2)]}{p^4}}{\left(\frac{rq}{P^2}\right)^2}$$

$$= \frac{rq[P^2 + 3q(r+2)]}{p^4}\left(\frac{P^2}{rq}\right)^2$$

$$\beta_2 = \frac{[P^2 + 3q(r+2)]}{rq}$$

Truncated Distribution

Let X be a random variable with p.d.f. (or p.m.f.) f(x). The distribution of X is said. to be truncated at the point X = a if all the values of X ≥ a are discarded. Hence the p.d.f. (or p.m.f.) g (.) of the distribution, truncated at X = a is given by:

$$g(x) = \frac{f(x)}{p(x > a)}; x > a$$

$g(x) = \frac{f(x)}{\sum f(x)}; x > a$ (for discrete random variable)

$g(x) = \frac{f(x)}{\int_a^\infty f(x)dx}; x > a$(for continuous random variable)

For the continuous random variable x the rth moment (about origin) for the truncated distribution is given by

$$\mu_r' = E(x^r)$$
$$= \int_a^{\infty} x^r\, g(x)dx$$

$$= \frac{\int_a^{\infty} x^r\, g(x)dx}{\left(\int_a^{\infty} f(x)\, dx\right)}$$

Truncated Binomial Distribution

Let f(x) be the p.m.f. of X - B (n. p) variate. Then the p.m.f. g (x) of the Binomial distribution truncated at X = 0 is given by

$f(x) = n_{C_x} p^x q^{n-x}\ ; x = 0,1,2, \dots\dots n;$ q = 1- P.

$$g(x) = \frac{f(x)}{p(x > a)}; x > a = \frac{f(x)}{1 - p(x = 0)}$$

$$= \frac{f(x)}{1 - f(0)} = \frac{n_{C_x} p^x q^{n-x}}{1 - n_{C_0} p^0 q^{n-0}}$$

$$g(x)\ = \frac{n_{C_x} p^x q^{n-x}}{1 - q^n}; x = 1,2, \dots . n$$

Mean and Variance of Truncated Binomial Distribution

Mean

$$g(x)\ = \frac{n_{C_x} p^x q^{n-x}}{1 - q^n}; x = 1,2, \dots . n$$

$$\mathrm{E(X)} = \sum_{x=1}^{n} x\ g(x)$$

$$= \sum_{x=1}^{n} x\ \frac{n_{C_x} p^x q^{n-x}}{1 - q^n}$$

$$= \frac{1}{1 - q^n} \sum_{x=1}^{n} x\ n_{C_x} p^{x-1} p q^{n-x-1+1}$$

$$= \frac{1}{1 - q^n} p \sum_{x=1}^{n} x\ n_{C_x} p^{x-1} q^{n-1-(x+1)}$$

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

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

$$= \frac{1}{1-q^n} np \sum_{x=1}^{n} n-1_{C_{x-1}} p^{x-1} q^{n-1}$$

$$= \frac{1}{1-q^n} np \left[n-1_{C_0} p^0 q^{n-1} + \cdots + n-1_{C_{n-1}} p^{n-1} q^0\right]$$

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

$$E(x) = \frac{np}{1-q^n}$$

Variance

$Var(x) = E(x^2) - [E(x)]^2$

$$E(x^2) = \sum_{x=1}^{n} x^2 \ g(x)$$

$$= \sum_{x=1}^{n} x^2 \ \frac{n_{C_x} p^x q^{n-x}}{1-q^n}$$

$$= \frac{1}{1-q^n} \sum_{x=1}^{n} x^2 \ n_{C_x} p^x q^{n-x}$$

$$= \frac{1}{1-q^n} \left[\sum_{x=0}^{n} x^2 \ n_{C_x} p^x q^{n-x} - 0\right]$$

$$= \frac{1}{1-q^n} [npq + n^2p^2]$$

$$var(x) = \frac{1}{1-q^n} [npq + n^2p^2] - \left(\frac{np}{1-q^n}\right)^2$$

$$= \frac{1}{1-q^n} \left[npq + n^2p^2 - \frac{n^2p^2}{1-q^n}\right]$$

Truncated Poisson Distribution

Let f(x) be the p.m.f. of X - P(λ) variate. Then the p.m.f. g (x) of the Poisson distribution truncated at X = 0 is given by

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}\ ; x = 0,1,2, \dots , n$$

$$g(x) = \frac{f(x)}{p(x > a)}; x > a = \frac{f(x)}{1 - p(x = 0)}$$

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

$$g(x)\ = \frac{e^{-\lambda}\lambda^{x}}{x!(1 - e^{-\lambda})}; x = 0,1,2, \dots , n$$

Mean and Variance of Truncated Poisson Distribution

Mean

$$g(x)\ = \frac{e^{-\lambda}\lambda^{x}}{x!(1 - e^{-\lambda})}; x = 0,1,2, \dots , n$$

$$E(X) = \sum_{x=0}^{n} x\ g(x)$$

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

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

$$= \frac{1}{1 - e^{-\lambda}} . \lambda$$

$$E(X) = \frac{\lambda}{1 - e^{-\lambda}}$$

Variance

Var(x) = E(x^2) –[E(x)]^2

$$E(X)^2 = \sum_{x=0}^{n} x^2\ g(x)$$

$$= \sum_{x=0}^{n} x(x - 1) + x\ g(x)$$

$$= \sum_{x=0}^{n} x(x-1)g(x) + \sum_{x=0}^{n} x\, g(x)$$

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

$$= \frac{1}{1-e^{-\lambda}}\sum_{x=0}^{n} x(x-1)\frac{e^{-\lambda}\lambda^x}{x(x-1)(x-2)!} + \frac{\lambda}{1-e^{-\lambda}}$$

$$E(X)^2 = \frac{1}{1-e^{-\lambda}}[\lambda^2 + \lambda]$$

$\mathrm{Var}(x) = E(x^2) - [E(x)]^2$

$$= \frac{1}{1-e^{-\lambda}}[\lambda^2 + \lambda] - \left(\frac{\lambda}{1-e^{-\lambda}}\right)^2$$

$$= \frac{1}{1-e^{-\lambda}}\left[\lambda^2 + \lambda - \frac{\lambda^2}{1-e^{-\lambda}}\right]$$

$$var(X) = \frac{\lambda}{1-e^{-\lambda}}\left[\lambda + 1 - \frac{\lambda}{1-e^{-\lambda}}\right]$$

Power Series Distribution

A discrete random variable X is said to follow a generalized power series distribution (g.p.s.d) if its probability mass function is given by

$$p(X = x) = \frac{a_x\theta^x}{f(\theta)};\ x = 0,1,2,\ldots.,a_x \geq 0$$

Where $f(\theta)$ is a generating function

(i.e,) $f(\theta) = \sum_{x \in s} a_x\theta^x \,, \theta \geq 0$

So that $f(\theta)$ is positively finite and differentiable and S is non empty countable subset of non negative integers.

Moment Generating Function of Power Series Distribution

$$p(X = x) = \frac{a_x\theta^x}{f(\theta)};$$

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

$$= \sum_{x=0}^{\infty} e^{tx}\, p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} \frac{a_x \theta^x}{f(\theta)}$$

$$= \frac{1}{f(\theta)} \sum_{x=0}^{\infty} e^{tx} a_x \theta^x$$

$$= \frac{1}{f(\theta)} \sum_{x=0}^{\infty} a_x (\theta e^t)^x \left(\sum_{x=0}^{\infty} \frac{a_x \theta^x}{f(\theta)} = 1;\ \sum_{x=0}^{\infty} a_x \theta^x = f(\theta) \right)$$

$$M_x(t) = \frac{1}{f(\theta)} f(\theta e^t)$$

Cumulates of Power Series Distribution

$$k_x(t) = \log M_x(t)$$

$$= \log \left[\frac{f(\theta e^t)}{f(\theta)} \right]$$

$$\sum_{x=0}^{\infty} k_r \left(\frac{t^r}{r!} \right) = log\ f(\theta e^t) - log f(\theta) \rightarrow (1)$$

Diff w.r.t θ and t

$\sum_{x=0}^{\infty} \frac{\partial}{\partial \theta} k_r \left(\frac{t^r}{r!} \right) = e^t \frac{f'(\theta e^t)}{f(\theta e^t)} - \frac{f'(\theta)}{f(\theta)} \rightarrow (2)$ and

$$\sum_{r=1}^{\infty} k_r \left(\frac{r t^{r-1}}{r!} \right) = \frac{\theta e^t f'(\theta e^t)}{f(\theta e^t)} \rightarrow (3)$$

Subtracting (3) from θ times (2), we get

$$\theta \sum_{r=1}^{\infty} \frac{\partial}{\partial \theta} k_r \left(\frac{t^r}{r!} \right) = \frac{\theta e^t f'(\theta e^t)}{f(\theta e^t)} - \frac{\theta f'(\theta)}{f(\theta)}$$

$$= \sum_{r=1}^{\infty} k_r \left(\frac{r t^{r-1}}{r!} \right) - \frac{\theta f'(\theta)}{f(\theta)}$$

Comparing the powers of t on both sides, we get

$$0 = k_1 - \frac{\theta f'(\theta)}{f(\theta)}$$

$$k_1 = \frac{\theta f'(\theta)}{f(\theta)}$$

$$k_{r+1} = \theta \frac{d}{d\theta} k_r; r = 1.2 \ldots\ldots (comparing\ co\ efficient\ of \frac{t^r}{r!})$$

$$mean \ = k_1 = \frac{\theta f^{'}(\theta)}{f(\theta)}$$

Logarithmic Series Distribution

Let $f(\theta) = -\log(1-\theta)$

$$f(\theta) = \sum_{x \in s} a_x \theta^x \ , s = 0,1,2, \dots$$

$$-\log(1-\theta) = \sum_{x=0}^{\infty} a_x \theta^x$$

$$-\left(-\frac{\theta}{1} - \frac{\theta^2}{2} - \frac{\theta^3}{3}\right) = a_1\theta^1 + a_2\theta^2 + \cdots$$

$$\frac{\theta}{1} + \frac{\theta^2}{2} + \frac{\theta^3}{3} + \cdots \dots . = a_1\theta^1 + a_2\theta^2 + \cdots$$

$$a_1 = 1$$

$$a_2 = 2$$

$$a_3 = 3$$

$$a_x = \frac{1}{x}$$

Now

$$p(X = x) = \frac{a_x \theta^x}{f(\theta)}$$

$$= \frac{1}{x} \frac{\theta^x}{-\log(1-\theta)} ; x = 1,2, \dots .$$

Which is the p.m.f of Logarithmic series distribution with parameter θ.