

Distribution Theory

Department of Statistics & Informative

Master Students (Second Semester)

Academic Year (2023-2024)

Assistant Professor

Dr. Luceen Immanuel Kework

References of Distribution Theory

- 1) Introduction to Theory of Statistics. By Mood Graybill & Boes, 3rd edition (1974).**
- 2) Introduction to the Theory of Statistics. By Mood Graybill, 2nd edition (1963).**
- 3) Theoretical Statistics. By Cox D.R. & Hinkley D.V. (1974).**
- 4) Introduction to Mathematical Statistics. By Hogg & Craig, 4th edition (1978).**
- 5) Introduction to Mathematical Statistics. By Hogg Mekean & Craig, 6th edition (2006).**
- 6) Statistical Theory. By Lindgren B. (1982).**
- 7) An Introduction to Probability & Statistics. By Rohatgi Vijary & Saleh, 2nd edition (2001).**
- 8) Mathematical Statistics. By Freund Walpole & Meyer, (1979).**
- 9) Introductory to Probability & Statistics Applications. By Paul Meyer, (1978).**
- 10) Probability & Statistics for Engineers & Scientists. By Walpole Myers Myers Ye 7th edition (2002).**
- 11) Modern Probability Theory and its Applications. Emmanuel Parzen.**
- 12) Probability Theory. By Filler.**

Statistical Distributions

1) Discrete Uniform Distribution

Suppose X is a discrete uniform r.v., over the integer set $\{a, a+1, \dots, b\}$.

Assumptions:

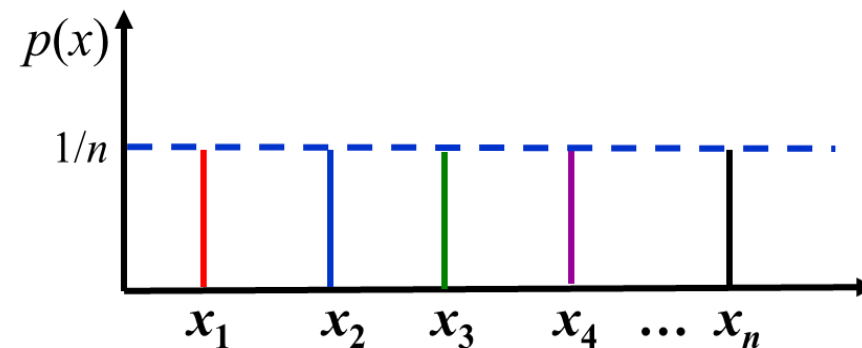
1) $X \in \{a, a+1, \dots, b\}$

2) The values in the range are **equally likely**, i.e., each of n values in the range of r.v. X has probability $\frac{1}{n}$ (all outcomes have equal probability). $X \sim (a, b)$ Probability Mass Function:

$$p.m.f = p(x) = p(X = x) = \frac{1}{b - a + 1}, x = a, a + 1, \dots, b$$

Let; $n = b - a + 1$

$$\therefore p(x) = \frac{1}{n}, x = a, a + 1, \dots, b$$



The p.m.f of uniform distribution

In a special case, a r.v is defined to have a discrete uniform distribution. If the p.m.f. of X is given by:

$$p(x) = \begin{cases} \frac{1}{n} & x=1,2,\dots,n \\ 0 & o.w \end{cases}$$

Where the parameter (n) is positive integer: $X \sim D.U(1, n)$. Each point of X has the same probability of appearing.

Properties of Uniform Distribution

1) If X has a uniform dist., then the c.d.f of X

F(x) is defined as;

$$F(x) = p(X \leq x) = \sum_{x=1}^x p(x) = \sum_{x=1}^x \frac{1}{n} = \frac{x}{n}$$

$$F(x) = \begin{cases} 0 & , x < 1 \\ 1/n & , 1 \leq x < 2 \\ 2/n & , 2 \leq x < 3 \\ \cdot & \\ \cdot & \\ \frac{n-1}{n} & , n-1 \leq x < n \\ 1 & , x \geq n \end{cases}$$

2) The mean and the variance of uniform distribution

$$\therefore E(X) = \frac{b - a + 1 + 2a - 1}{2} = \frac{a + b}{2} = \textit{Median}$$

$$\begin{aligned} \text{Variance} = v(X) &= (n + 1) \frac{b - a}{12} = \frac{(n + 1)(n - 1)}{12} \\ &= \frac{n^2 - 1}{12} = \frac{(b - a + 1)^2 - 1}{12} \end{aligned}$$

3) The moment generating function (m.g.f.) of uniform distribution

$$M_x(t) = E(e^{tX}) = \sum_{x=a}^b e^{tx} \frac{1}{n} = \frac{1}{n} \sum_{x=a}^b e^{tx}$$

$$\therefore M_x(t) = \frac{e^{at}}{n} \left(\frac{1 - e^{nt}}{1 - e^t} \right)$$

$$= \begin{cases} \frac{e^{at}}{n} \left(\frac{1 - e^{nt}}{1 - e^t} \right) & , \quad t \neq 0 \\ 1 & , \quad t = 0 \end{cases}$$

4) Mode of uniform distribution

$$\because \frac{f(m)}{f(m+1)} = 1 \text{ and } \frac{f(m)}{f(m-1)} = 1,$$

\because p(each of n values in the range of r.v. X) has probability = $\frac{1}{n}$ (equally likely)

\therefore on mode in this distribution, or all points in range is mode (multi – mode).

5) A median of a distribution of one r.v X of the continuous or discrete type, is the value of X that satisfies the following conditions:

$$\left. \begin{array}{l} 1- p(X < m) = \sum_{x < m} p(x) \leq \frac{1}{2} \\ 2- p(X \leq m) = \sum_{x \leq m} p(x) \geq \frac{1}{2} \end{array} \right\} \text{median} = \text{mean in D.U. dist.}$$

1. Bernoulli Distribution

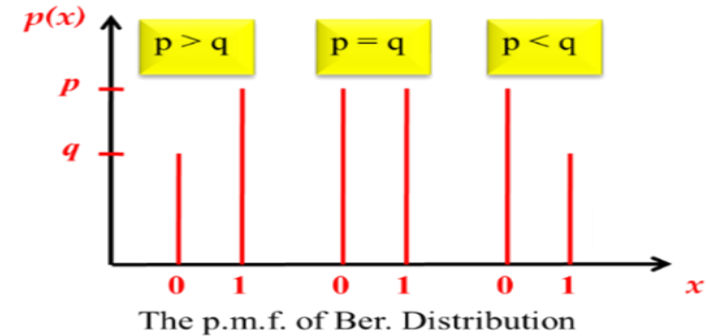
If a r.v. X in a single trial has only two outcomes, either ($x = 0$) or ($x = 1$), which are success and failure, or Yes and No, good and bad, defective and non-defectiveetc. , with;

$$p(\text{success}) = p \quad , \quad p(\text{failure}) = 1 - p = q$$

The p.m.f. of X ;

$$p(x; p) = p(x) = \begin{cases} p^x(1-p)^{1-x} & , \quad x = 0,1 \\ 0 & \text{o.w} \end{cases}$$

Where (p) is the parameter of distribution, that satisfies ($0 < p < 1$),



where;

p : probability of success, $p(S) = p(x = 1)$.

$1 - p = q$: probability of failure, $p(F) = p(x = 0)$

$(p + q) = p + (1 - p) = 1$.

$X \sim \text{Ber}(1, p)$,

$X \sim \text{Ber} \left(\begin{matrix} 1 \\ \text{No. of success} \\ \text{trials} \end{matrix}, \begin{matrix} p \\ \text{probability of} \\ \text{success} \end{matrix} \right)$

$$X \sim \text{Ber}(1, p) \Rightarrow p(x, p) = \begin{cases} p & , \quad x = 1 \\ 1 - p & , \quad x = 0 \\ 0 & \text{o.w} \end{cases}$$

Properties of Bernoulli Distribution

a) The mean and the variance of X

$$\text{mean} = E(X) = \sum_{x=0}^1 x p(x) = \sum_{x=0}^1 x p^x (1-p)^{1-x} = 0 + (1) p^1 q^0 = p$$

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_{x=0}^1 x^2 p^x (1-p)^{1-x} = 0 + (1)^2 p^1 q^0 = p$$

$$\therefore \text{var}(X) = p - p^2 = p(1-p) = pq$$

b) The cumulative dist. function (c.d.f.) of $X \sim \text{Ber}(1, p)$

$$F(x) = p(X \leq x) = \begin{cases} 0 & , x < 0 \\ q & , 0 \leq x < 1 \\ p + q = 1 & , x \geq 1 \end{cases}$$

c) The moment generating function (m.g.f.) when $X \sim \text{Ber}(1, p)$

$$\begin{aligned} M_X(t) &= E e^{tX} = \sum_{x=0}^1 e^{tx} p(x) = \sum_{x=0,1} e^{tx} p^x (1-p)^{1-x} \\ &= \sum_{x=0,1} (pe^t)^x (q)^{1-x} = (pe^t)^0 (q)^{1-0} + (pe^t)^1 (q)^{1-1} = q + pe^t \end{aligned}$$

$$\therefore M_X(t) = q + pe^t$$

$$M'_X(0) = \mu_X = pe^t \Big|_{t=0} = p$$

$$M''_X(0) = E(X^2) = pe^t \Big|_{t=0} = p$$

$$\therefore \text{var}(X) = M''_X(0) - [M'_X(0)]^2 = p - p^2 = p(1-p) = pq$$

The Mode

A mode is that value of X that gives the greatest probability. In other words, is the value of X that maximizes the p.d.f. , such that:-

- 1) If X is discrete r.v. , then we substitute values of X in p.m.f. $p(x)$ directly, then the value that gives the greatest probability is the mode of distribution.
- 2) If X is continuous r.v., then we must take the second derivative of p.d.f., then if this derivative is less than zero then the value of X is the mode of distribution.

d) The Mode of $X \sim \text{Ber}(1, p)$ (Mo)

$$Mo = x = \begin{cases} 0 & , q > p \\ 0,1 & , q = p \\ 1 & , q < p \end{cases}$$

The Median

A median of a distribution of one r.v X of the continuous or discrete type, is the value of X that satisfies the following conditions:

$$\left. \begin{array}{l} 1- p(X < m) = \sum_{x < m} p(x) \leq \frac{1}{2} \\ 2- p(X \leq m) = \sum_{x \leq m} p(x) \geq \frac{1}{2} \end{array} \right\} \text{ for a discrete r.v.}$$

$$3- F(m) = p(X \leq m) = \int_{-\infty}^m f(x) dx = \frac{1}{2} \leftarrow \text{for continuous r.v}$$

e) The Median of $X \sim \text{Ber}(1, p)$ (Me)

$$\because x = 0, 1 \quad , \quad Me = \frac{0 + 1}{2} = \frac{1}{2}$$

$\because \frac{1}{2}$ is not within range, \therefore Bernoulli distribution does not have median.

Related Distributions

- 1) Binomial Dist.
- 2) Geometric Dist.
- 3) Negative Binomial Dist.
- 4) Beta Dist.
- 5) Categorical Dist.

Estimation Theory for Bernoulli Distribution

Methods of Estimation

1. Moments Estimation Method

$$m_k = \frac{\sum X_i^k}{n} \quad M_k = E(X^k)$$

$$m_k = M_k$$

$$m_1 = \frac{\sum X_i}{n} = \bar{X}, \quad M_1 = E(X) = p$$

$M_1 = m_1 \rightarrow \therefore \hat{p} = \bar{X} \rightarrow$ Proportion of success is estimator for probability of success.

2. Maximum Likelihood Estimation

Properties of Maximum Likelihood Estimation

The m.l.e for Bernoulli parameter $X \sim \text{Ber}(p)$,

Let X_1, X_2, \dots, X_n denote a random sample from Bernoulli distⁿ $\text{Ber}(p)$, then the m.l.e for p .

$\because X \sim \text{Ber}(p)$

$$f(x; p) = p^x(1 - p)^{1-x}, \quad x = 0, 1$$

$\because X$'s are indep.

$$\begin{aligned} L(p) &= f(x_1, x_1, \dots, x_1; p) = \prod f(x_i; p) \\ &= p^{\sum x_i} (1 - p)^{n - \sum x_i} \end{aligned}$$

$$\ln L(p) = \sum x_i \ln(p) + (n - \sum x_i) \ln(1 - p)$$

$$\frac{\partial \ln L(p)}{\partial p} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p}, \quad \frac{\partial \ln L(p)}{\partial p} = 0$$

$$\frac{\sum x_i}{p^\wedge} - \frac{n - \sum x_i}{1 - p^\wedge} = 0$$

$$\frac{(1 - p^\wedge)\sum x_i - p^\wedge(n - \sum x_i)}{p^\wedge(1 - p^\wedge)} = 0$$

$$\Sigma x_i - p^{\wedge} \Sigma x_i - np^{\wedge} + p^{\wedge} \Sigma x_i = 0$$

$$\Sigma x_i - np^{\wedge} = 0$$

$$\Sigma x_i = np^{\wedge} \quad p^{\wedge}_{m.l.e} = \frac{\Sigma X_i}{n} = \bar{X}$$

$$\frac{\partial^2 \ln L(p)}{\partial p^2} = -\frac{\Sigma x_i}{p^2} - \frac{n - \Sigma x_i}{(1-p)^2} < 0$$

$\therefore p^{\wedge} = \bar{X}$ is m.l.e for $p \rightarrow$ proportion of success (which gives the greatest value .

$\hat{p}_{m.l.e} = \frac{\Sigma X_i}{n} = \bar{X} \rightarrow$ proportion of success is an estimator for probability of success

H.W: Find Bayes estimator for parameter of $\text{Ber}(p)$, using non informative and informative prior probability.

Properties of Estimator

1. Unbiasedness

If $E(\hat{p}) = p$, then an estimator is unbiased estimator for p .

$$E(\bar{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{n}{n} E(X) = p, X \sim \text{Ber}(p), E(X) = p$$

$\therefore \bar{X}$ is unbiased estimator for p .

2. Unbiased in Limit

An estimator $\hat{\theta}$ is said to be unbiased in limit for θ iff:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

3. Consistency (Limiting Property)

Any statistics that equal to the parameter or converges stochastically to the parameter θ is called consistent statistics.

A consistent estimator: That the estimator gets closer to the parameter value as n increases without limit.

Definition 1: An estimator $\hat{\theta}$ of the parameter θ of $f(x; \theta)$ is called consistent estimator for θ , iff;

$$\lim_{n \rightarrow \infty} p(|\hat{\theta} - \theta| < \varepsilon) = 1 \quad , \quad \forall \varepsilon > 0$$

or;

$$\lim_{n \rightarrow \infty} p(|\hat{\theta} - \theta| \geq \varepsilon) = 0$$

$$\left. \begin{array}{l} p(|\hat{\theta} - \theta| < \varepsilon) \geq 1 - \frac{v(\hat{\theta})}{\varepsilon^2} \\ p(|\hat{\theta} - \theta| \geq \varepsilon) < \frac{v(\hat{\theta})}{\varepsilon^2} \end{array} \right\} \rightarrow (\text{Chebycheve inequality})$$

Theorem: Let $\hat{\theta}$ be an unbiased estimator for population parameter θ of $f(x; \theta)$, then $\hat{\theta}$ is said to be consistent estimator for θ iff;

1. $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta \rightarrow$ unbiased in limit ,
 2. $\lim_{n \rightarrow \infty} v(\hat{\theta}) = 0 \rightarrow$ There is no error
- $\lim_{n \rightarrow \infty} E(\hat{p}) = p$ with probability

Mean Square Error

One way of measuring the accuracy of an estimator is via its mean square error. The mean square error of an estimator $\hat{\theta}$ is defined as:

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + b^2(\hat{\theta})$$

Proof:

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= E(\hat{\theta} - \theta \mp E(\hat{\theta}))^2 \\ &= E(\{\hat{\theta} - E(\hat{\theta})\} + \{E(\hat{\theta}) - \theta\})^2 \\ &= E(\hat{\theta} - E(\hat{\theta}))^2 + E(E(\hat{\theta}) - \theta)^2 + 2 E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) \\ &= Var(\hat{\theta}) + b^2(\hat{\theta}) + zero \\ \therefore MSE(\hat{\theta}) &= Var(\hat{\theta}) + b^2(\hat{\theta}) \end{aligned}$$

The Score Function

The score function is the partial derivative of \ln the function $f(x;\theta)$ with respect to the parameter θ , is defined as;

$$S(x;\theta) = \frac{\partial}{\partial \theta} \ln f(x;\theta) = \frac{1}{f(x;\theta)} \frac{\partial}{\partial \theta} f(x;\theta)$$

or; *Score fun* = $S(x; \theta) = \frac{\partial \ln L(\theta)}{\partial \theta}$

Properties

1) The mean of the score is zero, $E(S(X; \theta)) = \text{zero}$ (under Regularity Condition)

Proof:

$$\begin{aligned} E(S(X; \theta)) &= \int_{R_x} s(x; \theta) f(x; \theta) dx = \int_{R_x} \frac{1}{f(x; \theta)} \frac{\partial}{\partial \theta} f(x; \theta) f(x; \theta) dx \\ &= \int_{R_x} \frac{\partial}{\partial \theta} f(x; \theta) dx = \frac{\partial}{\partial \theta} \int_{R_x} f(x; \theta) dx = \frac{\partial}{\partial \theta} (1) = \text{zero} \end{aligned}$$

2) The variance of the score is known as the Fisher Information (F.I), which is measure the information in the sample \mathcal{S} about the parameter θ , and can be written as;

$$\begin{aligned} V(S(x; \theta)) &= E \left(S(x; \theta) - E(S(x; \theta)) \right)^2 \\ &= E(S(x; \theta))^2 - (E(S(x; \theta)))^2, \quad E(S) = 0 \\ &= E(S(x; \theta))^2 \rightarrow \text{Information} \end{aligned}$$

$$F.I = I(\theta) = E \left(\frac{\partial \ln L(\theta)}{\partial \theta} \right)^2$$

If Fisher Information multiply by (n) , we get; $n I(\theta) \rightarrow \text{F.I in a rss}(n)$

The Rao- Cramer Inequality

Let X_1, X_2, \dots, X_n be a rssn from a distⁿ with p.d.f. $f(x ; \theta)$, and let $T = u(X_1, X_2, \dots, X_n)$ be an unbiased estimator for $\phi(\theta)$, then the variance of T satisfies the inequality;

$$V(T) \geq \frac{(\phi'(\theta))^2}{n E \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2} = \frac{(\phi'(\theta))^2}{\text{Var}(S)} = \frac{(\phi'(\theta))^2}{-n E \left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right)}$$

Notes:

1) We do not use (n) in case using the likelihood function in law.

$$\begin{aligned} \text{a) } F.I = V(S) &= -n E \left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right) \\ \text{b) } F.I = V(S) &= n E \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 \end{aligned} \left. \vphantom{\begin{aligned} \text{a) } F.I = V(S) &= -n E \left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right) \\ \text{b) } F.I = V(S) &= n E \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 \end{aligned}} \right\} f(x; \theta) \text{ for single}$$

$$\begin{aligned} \text{c) } F.I = V(S) &= -E \left(\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \right) \\ \text{d) } F.I = V(S) &= E \left(\frac{\partial \ln L(\theta)}{\partial \theta} \right)^2 \end{aligned} \left. \vphantom{\begin{aligned} \text{c) } F.I = V(S) &= -E \left(\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \right) \\ \text{d) } F.I = V(S) &= E \left(\frac{\partial \ln L(\theta)}{\partial \theta} \right)^2 \end{aligned}} \right\} L(x; \theta)$$

2) In general case, if T is unbiased estimator for $\phi(\theta)$ of $f(x; \theta)$.

$\frac{(\phi'(\theta))^2}{V(S)}$ is called Rao-Cramer Lower Bound (RCLB) (Minimum variance bound (MVB))

3) Special case of Rao- Cramer Inequality: If T unbiased estimator for θ , $E(T) = \theta$;

$$\phi(\theta) = \theta \quad \rightarrow \quad \phi'(\theta) = 1$$

$$\therefore \left(RCLB = \frac{1}{V(S)} \right) \quad , V(Y) \geq \frac{1}{F.I} \quad , F.I = V(\text{Score Function} = S)$$

4) $\because X$'s are independent

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \text{Likelihood function}$$

5) As more information increases, the variance decreases, and when the variance decreases, the estimator becomes more efficient, but Fisher increases.

4) Efficiency

How do we know that the estimator is an efficient estimator?

Defⁿ: Let T be unbiased estimator for θ of $f(x_i; \theta)$ and The ratio of the RCLB to the actual variance of any unbiased estimator for θ is called the efficiency;

$$eff = \frac{RCLB}{V(T)} \quad , \quad 0 \leq eff \leq 1$$

if $eff = 1 \Rightarrow T$ is called efficient estimator for θ .

Defⁿ: Let T be an unbiased estimator for $\phi(\theta)$, then we say that T is an efficient estimator for θ iff; (The variance is as low as possible).

$$V(T) = RCLB$$

$$eff(T) = eff(\hat{p} = \bar{X} = \frac{\sum X_i}{n}) = \frac{\frac{1}{F.I}}{V(T)} = \frac{1}{V(\hat{p})}$$

for Bernoulli distribution $X \sim \text{Ber}(p)$

$$f(x; p) = p^x (1 - p)^{1-x}$$

$$\ln f(x; p) = x \ln p + (1 - x) \ln(1 - p)$$

$$\frac{\partial \ln f(x; p)}{\partial p} = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)}$$

$$\begin{aligned} \text{F.I. in a rsn} &= n \text{E} \left(\frac{\partial \ln f(x; p)}{\partial p} \right)^2 \\ &= n \frac{E(X-p)^2}{p^2(1-p)^2} \\ &= \frac{n p (1-p)}{p^2(1-p)^2} = \frac{n}{p(1-p)} \end{aligned}$$

$$\frac{1}{\text{F.I.}} = \frac{p(1-p)}{n}$$

$$\therefore \text{eff}(T = \hat{p}) = \frac{\frac{1}{\text{F.I.}}}{V(\hat{p})} = \frac{p(1-p)/n}{p(1-p)/n} = 1$$

$$\therefore \hat{p} = \frac{\sum X_i}{n} \text{ is an efficient est. for } p$$

5) Sufficiency

Sufficiency estimator is containing all the information in the data about the parameter θ .

(e.g. $\bar{X} = \frac{\sum x_i}{n}$) The information in $(\sum x_i)$ is the same as the information in (\bar{X}) .

First Method (Fisher Information)

Let X_1, X_2, \dots, X_n be a rssn from the distⁿ with p.d.f. $f(x; \theta)$, an estimator $\hat{\theta}$ is sufficient estimator for the parameter θ if the Fisher information in $\hat{\theta}$ is equal to the Fisher information in a rss(n).

$$\text{F.I in a rss}(n) = \text{F.I in } T = \hat{\theta}$$

For Bernoulli distribution. Let X_1, X_2, \dots, X_n be a rssn from Bernoulli distⁿ $\text{Ber}(\theta)$. Show that $\hat{\theta} = \sum X_i$ is sufficient estimator for the parameter θ .

$\therefore X \sim \text{Ber}(\theta)$

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x}, \quad x = 0, 1$$

$$\ln f(x; \theta) = x \ln(\theta) + (1 - x) \ln(1 - \theta)$$

$$\frac{\partial \ln f(x; \theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1 - x)}{(1 - \theta)}$$

$$\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} = \frac{-x}{\theta^2} - \frac{(1 - x)}{(1 - \theta)^2}$$

$$\begin{aligned} F.I &= - E \left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right) = \frac{E(X)}{\theta^2} + \frac{E(1 - X)}{(1 - \theta)^2} \\ &= \frac{\theta}{\theta^2} + \frac{(1 - \theta)}{(1 - \theta)^2} = \frac{1}{\theta} + \frac{1}{(1 - \theta)} = \frac{1 - \theta + \theta}{\theta(1 - \theta)} = \frac{1}{\theta(1 - \theta)} \end{aligned}$$

$$n I(\theta) = - n E \left(\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right) = \frac{n}{\theta(1 - \theta)} \text{ is F.I. in a rss}(n)$$

$$\hat{\theta} = \sum X_i = x_1 + x_2 + \dots + x_n$$

$$X \sim \text{Ber}(\theta) \quad \Rightarrow \quad \sum X_i \sim \text{Bin}(n, \theta)$$

$$f(\sum x_i; \theta) = C_{\sum x_i}^n \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

$$\ln f(\sum x_i; \theta) = \ln C_{\sum x_i}^n + \sum x_i \ln(\theta) + (n - \sum x_i) \ln(1 - \theta)$$

$$\frac{\partial \ln f(\sum x_i; \theta)}{\partial \theta} = \text{zero} + \frac{\sum x_i}{\theta} - \frac{(n - \sum x_i)}{(1 - \theta)}$$

$$\frac{\partial^2 \ln f(\sum x_i; \theta)}{\partial \theta^2} = - \frac{\sum x_i}{\theta^2} - \frac{(n - \sum x_i)}{(1 - \theta)^2}$$

$$- E \left(\frac{\partial^2 \ln f(\sum x_i; \theta)}{\partial \theta^2} \right) = \frac{n E(X)}{\theta^2} + \frac{(n - n E(X))}{(1 - \theta)^2}$$

$$= \frac{n}{\theta} + \frac{n}{(1 - \theta)} = \frac{n}{\theta(1 - \theta)}$$

$$F.I \text{ in a rsn} = F.I \text{ in } \hat{\theta} = \sum X_i$$

$\therefore \hat{\theta} = \sum X_i$ is suff est for θ

Second Method (Conditional)

Let X_1, X_2, \dots, X_n be a r.s.s.n from the distⁿ with p.d.f. $f(x; \theta)$, and $\hat{\theta}$ be an estimator for θ , an estimator $\hat{\theta}$ is sufficient estimator for the parameter θ if the conditional p.d.f. of (X_1, X_2, \dots, X_n) given $\hat{\theta}$ does not contain the parameter θ :

$$f(x_1, x_2, \dots, x_n | \hat{\theta}) = \frac{f(x_1, x_2, \dots, x_n; \theta)}{g(\hat{\theta})}$$

Conditional p.d.f
doesn't depend on θ

Note: If the range depends on the parameter, in this case we can't find F.I; therefore, we use the second method (Conditional).

Third Method: Factorization Theorem

Let $\hat{\theta}$ be an estimator for the parameter of $f(x; \theta)$ such that the range does not depend on θ . Then the necessary and sufficient condition for an estimator $\hat{\theta}$ to be sufficient estimator, if there are two non-negative functions, such that:

$$f(x_1, x_2, \dots, x_n; \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \dots, x_n)$$

Theorem:

Let $\hat{\theta}$ be sufficient estimator for the parameter θ , and $u(\hat{\theta})$ be a one-to-one transformation, then $u(\hat{\theta})$ is sufficient estimator for θ .

Note: 1) \bar{X} is one to one transformation to $\sum X_i \cdot \Rightarrow \sum X_i = n \bar{X}$.

2) If we have more than one parameter, we use factorization theorem (third method) for sufficiency.

for Bernoulli distribution

$$\because X \sim \text{Ber}(\theta)$$

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x}, \quad x = 0, 1$$

$$\begin{aligned} f(x_1, x_2, \dots, x_n; \theta) &= \prod_{i=1}^n f(x_i; \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \quad \left. \vphantom{\prod_{i=1}^n} \right\} \times \frac{C_{\sum x_i}^n}{C_{\sum x_i}^n} \\ &= C_{\sum x_i}^n \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \times \frac{1}{C_{\sum x_i}^n}, \quad \text{free of } \theta \\ &= g(\hat{\theta} = \sum x_i; \theta) \times H(x) \quad \Rightarrow \therefore \hat{\theta} = \sum x_i \text{ is suff est for } \theta \end{aligned}$$

In case $T = \hat{\theta} = \bar{X}$ then we make up $n\bar{X}$ instead of $\sum X_i$, because $n\bar{X} = \sum X_i$ in any method.

Exponential Family

Let X has a p.d.f. $f(x; \theta)$, then the family of $f(x; \theta)$ is belong to exponential class of distribution iff it can be expressed as:

$$\begin{aligned} f(x; \theta) &= \exp(\ln f(x; \theta)) \\ &= \exp(A(\theta) B(x) + C(\theta) + S(x)) \quad \dots (1) \end{aligned}$$

$$\text{or;} \quad = g(\theta) h(x) e^{A(\theta) B(x)} \quad \dots (2)$$

Such that: $A(\theta) B(x)$ must have to be for exponential class. Then a r.v. is said to have exponential family.

Q// Is Bernoulli distribution belongs to exponential family?

$$\begin{aligned}f(x; \theta) &= \theta^x (1 - \theta)^{1-x} \\&= \exp(x \ln(\theta) + (1 - x) \ln(1 - \theta)) \\&= \exp\left(x \ln\left(\frac{\theta}{1 - \theta}\right) + \ln(1 - \theta)\right)\end{aligned}$$

$$A(\theta) = \ln \frac{\theta}{1 - \theta}, \quad B(x) = x, \quad C(\theta) = \ln(1 - \theta), \quad D(x) = 0$$

\therefore Bernoulli distribution belongs to exponential family

2. Binomial Distribution

In (n) trials, let the probability of an event occurring in each trial be equal to (p) , and let all trials be independent, then the total number of success in (n) independent Bernoulli trials is a r.v. X having a binomial distribution with p.m.f. is given by;

$$f(x; n, p) = p(X = x) = \begin{cases} C_x^n p^x q^{n-x} & , x = 0, 1, 2, \dots, n \\ 0 & o.w \end{cases}$$

where (n) and (p) are positive parameters, such that $(0 \leq p \leq 1)$, n : No. of trials (positive integer), i.e., $X \sim b(n, p)$.

C_x^n : A combination is the number of ways to choose a sample of x elements from a set of n distinct objects where order does not matter.

x : the number of successful trials.

p : probability of success in a single trial.

$n - x$: Number of failures in (n) trials.

q : probability of failure. $p + q = 1$

Remarks:

a) The binomial dist. reduces to Bernoulli dist. if $(n = 1)$.

b) Binomial experiment it must have four properties:

1. There must be a fixed number of trials.
2. Each trial must have two possible outcomes.
3. All trials must have the same probability of success.
4. The trials must be independent of each other.

Properties of the Binomial Distribution

1) Let X be a r.v. with $X \sim b(n, p)$, then; The mean and the variance of X .

$$\begin{aligned} \text{mean} = E(X) &= \sum_{x=0}^n x p(x) = \sum_{x=0}^n x C_x^n p^x q^{n-x} = \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\ &= \sum_{x=0}^n x \frac{n(n-1)!}{x(x-1)!(n-x)!} p p^{x-1} q^{n-x} = n p \sum_{x=0}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \end{aligned}$$

Let;

$$\left. \begin{array}{l} y = x - 1, \quad m = n - 1 \\ x = y + 1, \quad n = m + 1 \end{array} \right\} \text{Range: } \begin{cases} x = 0 \Rightarrow y = -1, \quad x = n \Rightarrow y = n - 1 \\ \therefore y = -1, 0, 1, 2, \dots, n - 1 \end{cases}$$

$$\therefore \text{mean} = n p \underbrace{\sum_{y=-1}^{n-1} \frac{m!}{y!(m-y)!} p^y q^{m-y}}_{=1} = n p$$

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

$$\therefore E(X^2) = n(n-1)p^2 \underbrace{\sum_{y=-2}^{n-2} \frac{m!}{y!(m-y)!} p^y q^{m-y}}_{=1} + np = n(n-1)p^2 + np$$

$$\text{var}(X) = n(n-1)p^2 + np - [np]^2 = n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p)$$

$$\therefore \text{var}(X) = npq$$

2) The moment generating function (m.g.f.) of $X \sim \text{bin}(n, p)$

$$M_X(t) = E e^{tX} = \sum_{x=0}^n e^{tx} p(x) = \sum_{x=0}^n e^{tx} C_x^n p^x q^{n-x} = \sum_{x=0}^n C_x^n (pe^t)^x q^{n-x}$$

$$\therefore (a+b)^n = \sum_{x=0}^n C_x^n a^x b^{n-x} \quad \text{by law}$$

$$\Rightarrow a = pe^t, \quad b = q$$

$$\therefore M_X(t) = (pe^t + q)^n \Big|_{t=0} = (p+q)^n = 1^n = 1$$

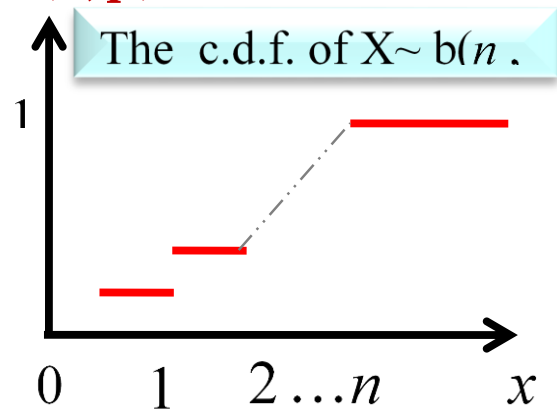
$$\therefore M_X(t) = (pe^t + q)^n$$

To find the mean and the variance by the m.g.f. of X .

$$M_X(t) = E e^{tX} = (pe^t + q)^n$$

3) The Cumulative Dist. Function (C.D.F.) of $X \sim b(n, p)$.

$$F(x) = p(X \leq x) = \begin{cases} 0 & , x < 0 \\ \sum_{u=0}^x C_u^n p^u q^{n-u} & , 0 \leq x < n \\ 1 & , x \geq n \end{cases}$$



4) Addition Property

If $X \sim b(n, p)$, then $Y = \sum_{i=1}^n X_i$ distributed Binomial distribution $b(\sum_{i=1}^n n_i, p)$.

Proof:

$$M_Y(t) = E e^{tY} = E e^{t \sum X_i} = E e^{\sum t X_i} = E(e^{t X_1} e^{t X_2} \dots e^{t X_n})$$

$$= \prod_{i=1}^n E e^{t X_i} = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (q + p e^t)^{n_i} = (q + p e^t)^{\sum_{i=1}^n n_i}$$

For example; $Y = \underset{\substack{\downarrow \\ b(n_1, p)}}{X_1} + \underset{\substack{\downarrow \\ b(n_2, p)}}{X_2} + \underset{\substack{\downarrow \\ b(n_3, p)}}{X_3} \quad \therefore Y = \sum_{i=1}^n X_i \sim b(\sum n_i, p)$

$\rightarrow Y \sim b(n_1 + n_2 + n_3, p)$

5) The mode of Binomial Distribution

If the mode of distribution of X is unique?

$$\text{Mode of } X(\text{Bin}) = m = [(n + 1)p]$$

Mode[Y] = greater integer less than or equal.

If the mode is an integer, it has two modes.

However, if the mode contains a fraction, then it has one mode

Let: $\varepsilon = \text{fraction} \rightarrow \text{then}; m = (n + 1)p + \varepsilon$

If $\varepsilon = 0$ then mode = $(m, m - 1) \rightarrow [(n + 1)p, (n + 1)p - 1]$

Mode integer = (m_1, m_2)

Mode fraction = (m)

Ex: $x = 6.1 \rightarrow \text{mode} = 6$

$x = 6.99 \rightarrow \text{mode} = 6$

if $x = 6$ (integer) $\rightarrow \text{mode} = 6, 5$

Ex: Let; $p = 0.7, n = 5$

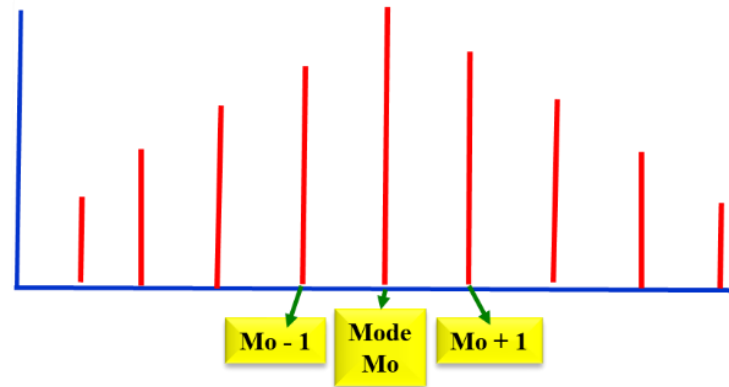
$$m = (n + 1)p = 6(0.7) = 4.2 \rightarrow \varepsilon = 0.2$$

$$\therefore m = 4$$

Ex: Let; $p = 0.3, n = 19$

$$m = (n + 1)p = 6$$

$$\therefore m = 6, 5$$



5) The median of Binomial Distribution

The median of Bin. Dist. is; $me = [np]$ if an integer (when me is an integer)

Proof: We have the following empirical relation between mean, mode and median:

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$$

Or; $\mu - mo = 3(\mu - me)$

In this case, $mo = (n + 1)p - \varepsilon$, $\rightarrow \varepsilon = 0$

$$me = \mu - \left(\frac{\mu - mo}{3}\right)$$

$$= \frac{2\mu + mo}{3} \rightarrow = \frac{2\bar{X} + mo}{3} \text{ for sample}$$

Take the case (np) is an integer;

$$me = \frac{2np + (n+1)p}{3} = \frac{2np + np + p}{3} = np + \frac{p}{3} \cong np$$

is an integer fraction

3. Poisson: The binomial distribution converges towards the [Poisson distribution](#) as the number of trials goes to infinity while the product np converges to a finite limit. Therefore, the Poisson distribution with parameter $\lambda = np$ can be used as an approximation to $B(n, p)$ of the binomial distribution if n is sufficiently large and p is sufficiently small. When p is very small and (n) large then the $\text{Bin}(n, p)$ distribution $\rightarrow \lambda = np$ become Poisson distribution, $\text{Bin}(n, p) \sim \text{Poi}(np)$.

According to rules of thumb, this approximation is good if $n \geq 20$ and $p \leq 0.05$ such that $np \leq 1$, or if $n > 50$ and $p < 0.1$ such that $np < 5$, or if $n \geq 100$ and $np \leq 10$.

Ex: $p = 0.01$, $n = \binom{200}{1000} \rightarrow \lambda = np = 2$ or 10 .

4. Beta: The binomial distribution and beta distribution are different views of the same model of repeated Bernoulli trials. The binomial distribution is the PMF of k successes given n independent events each with a probability p of success. Mathematically, when $\alpha = k + 1$ and $\beta = n - k + 1$, the beta distribution and the binomial distribution are related by a factor of $n + 1$:

$$\text{Beta}(p; \alpha, \beta) = (n + 1)\text{Bin}(k; n, p)$$

Beta distributions also provide a family of prior probability distributions for binomial distributions in Bayesian inference.

Q// Is Binomial distribution belongs to exponential family?

$$f(x; \theta, n) = \binom{n}{x} p^x (1 - \theta)^{n-x}$$

$$\exp(\ln f(x; \theta, n)) = \exp\left(\ln \binom{n}{x} + x \ln(\theta) + (n-x) \ln(1-\theta)\right)$$

$$= \exp\left(x \ln\left(\frac{\theta}{1-\theta}\right) + n \ln(1-\theta) + \ln \binom{n}{x}\right)$$

$$A(\theta) = \ln \frac{\theta}{1-\theta}, \quad B(x) = x = \sum_{i=1}^n x_i$$

$$, C(\theta) = \ln(1-\theta), \quad D(x) = \ln \binom{n}{x}$$

\therefore Binomial distribution belongs to exponential family.

Let X_1, X_2, \dots, X_n be a rssi from Binomial distⁿ $\text{Bin}(m, \theta)$, find the m.l.e for θ .

$$X \sim \text{Bin}(m, \theta)$$

$$f(x; \theta) = C_x^m \theta^x (1-\theta)^{m-x}, \quad x = 0, 1, \dots, m$$

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n C_{x_i}^m \theta^{\sum x_i} (1-\theta)^{\sum (m-x_i)}$$

$$\ln L(\theta) = \ln \prod_{i=1}^n C_{x_i}^m + \sum x_i \ln(\theta) + \sum (m-x_i) \ln(1-\theta)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \text{zero} + \frac{\sum x_i}{\theta} + \frac{\sum (m-x_i)}{(1-\theta)} \times (-1) = \frac{\sum x_i}{\theta} - \frac{\sum (m-x_i)}{(1-\theta)}$$

$$= \frac{\sum x_i(1-\theta) - \theta \sum (m - x_i)}{\theta(1-\theta)} = \frac{\sum x_i - \theta \sum x_i - nm\theta + \theta \sum x_i}{\theta(1-\theta)} = \frac{\sum x_i - nm\theta}{\theta(1-\theta)}$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0$$

$$\frac{\sum x_i - nm\theta}{\theta(1-\theta)} = 0 \Rightarrow \sum x_i - nm\theta = 0 \Rightarrow \sum x_i = nm\theta \Rightarrow \hat{\theta} = \frac{\sum X_i}{nm}$$

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} = \frac{-\sum x_i}{\theta^2} - \frac{\sum (m - x_i)}{(1-\theta)^2} = -\frac{\sum x_i}{\theta^2} - \frac{nm - \sum x_i}{(1-\theta)^2} < 0$$

$\therefore \hat{\theta}$ is m.l.e for θ .

H.W: Find estimator \hat{p} for $\text{Bin}(m, \theta)$ by using moments method.

Q// $X \sim \text{Bin}(m, \theta)$. Is $\hat{\theta}$ consistent estimator for θ

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \lim_{n \rightarrow \infty} E\left(\frac{\sum_{i=1}^n X_i}{nm}\right) = \lim_{n \rightarrow \infty} \frac{n E(X)}{nm} = \lim_{n \rightarrow \infty} \frac{nm\theta}{nm} = \theta$$

$$\lim_{n \rightarrow \infty} v(\hat{\theta}) = \lim_{n \rightarrow \infty} v\left(\frac{\sum_{i=1}^n X_i}{nm}\right) = \lim_{n \rightarrow \infty} \frac{nm\theta(1-\theta)}{n^2m^2} = \lim_{n \rightarrow \infty} \frac{\theta(1-\theta)}{nm} = 0$$

, $\therefore \hat{\theta}$ is consistent est. for θ .

Let X_1, X_2, \dots, X_n be a r.s.s.n from Binomial distⁿ $\text{Bin}(m, \theta)$, if $T = \hat{\theta} = \frac{\sum_{i=1}^n X_i}{n m}$ is an efficient estimator for $\phi(\theta) = \theta$.

Completeness

Completeness means uniqueness of unbiased estimators. This means if we have more than one unbiased estimators for θ then these all are the same if the distribution is complete.

Def: Let $f(x; \theta)$ denote a family of probability density function, **let $u(x)$ be a continuous function of (X) , then if $[E\{u(X)} = 0]$ implies $(u(x) = 0)$ at each point of (X)** , we say that the family of p.d.f. is complete.

Ex: assume that $[f(x; \theta)]$ is complete and let $u_1(x)$ and $u_2(x)$ be unbiased estimators for θ .

$$\text{i.e., } E[u_1(X)] = \theta$$

$$\text{and } E[u_2(X)] = \theta$$

$$E[u_1(X)] - E[u_2(X)] = 0$$

$$E[u_1(X) - u_2(X)] = 0$$

$$E[u(X)] = 0 \rightarrow u(x) = 0, \forall x$$

Because the distribution of X is complete.

$$u_1(x) - u_2(x) = 0$$

$$\rightarrow u_1(x) = u_2(x)$$

$$\begin{aligned} \text{Only } x &= 0 \\ E(X) &= E(0) = 0 \end{aligned}$$

Ex: Let X be a random variable from Binomial distribution $X \sim \text{Bin}(n, \theta)$. Show that the family of X is complete.

Sol:

$\because X \sim \text{Bin}(n, \theta)$

$$f(x; \theta, n) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

Let $u(x)$ be a continuous function of X , then;

$$E[u(X)] = 0 \quad \text{for all } \theta \in \Omega$$

$$E[u(Y)] = \sum_{x=0}^n u(x) \binom{n}{x} \theta^x (1 - \theta)^{n-x} = 0$$

$$\rightarrow u(0) \theta^0 (1 - \theta)^{n-0} + u(1) n \theta^1 (1 - \theta)^{n-1} + \dots + u(n) \theta^n (1 - \theta)^{n-n} = 0$$

$$\rightarrow u(0) (1 - \theta)^n + u(1) n \theta^1 (1 - \theta)^{n-1} + \dots + u(n) \theta^n = 0$$

$$\because \theta > 0 \rightarrow \theta \neq 0, n \neq 0$$

$$\because u(0) = u(1) = \dots = u(n) = 0 \quad \rightarrow u(x) = 0 \quad \forall x$$

$\therefore f(x; \theta, n)$ is complete

4) Geometric Distribution

Independent Bernoulli trials are performed (repeated) until the first success appears (occurs).

What is the distribution of the number of failures until the first success is observed?

Let the random variable X be the number of failures before the first success is observed. Since the first success may occur on the first trial, or second trial or third trial, and so on, X is a random variable with range space $\{0,1,2,3,\dots\}$ or $\{1,2,3,\dots\}$ with no (theoretical) upper limit. Then a r.v. X is defined to have geometric distribution if the p.m.f. of X given by;

The first form:

$$p(x) = p(x; p) = p(X = x) = \begin{cases} p q^x & , \quad x = 0, 1, 2, \dots \\ 0 & \text{o.w} \end{cases}$$

The second form:

$$p(x) = p(x; p) = p(X = x) = \begin{cases} p q^{x-1} & , \quad x = 1, 2, \dots \\ 0 & \text{o.w} \end{cases}$$

Where; the parameter (p) satisfies ($0 \leq p \leq 1$).

X : the number of failed trials before getting the first success.

Examples:

- 1) Tossing a coin continuously until the first success (head) appears.
- 2) Tossing a die continuously until the first success (number 4) appears.
- 3) Drawing objects from a box respectively with replacement until getting the defective object.

Properties of the Geometric Distribution

1) The mean and the variance of $X \sim \text{Geo}(p)$.

$$1) \text{ mean} = E(X) = \sum_{x=0}^{\infty} x p q^x = p \sum_{x=0}^{\infty} x q^x = p \sum_{x=0}^{\infty} x q^{x-1+1} = p q \sum_{x=0}^{\infty} x q^{x-1}$$

$$= p q \sum_{x=0}^{\infty} \frac{\partial}{\partial q} q^x \quad , \text{ because } \frac{\partial}{\partial q} q^x = x q^{x-1}$$

$$= p q \frac{\partial}{\partial q} \sum_{x=0}^{\infty} q^x \quad , \text{ by law } \sum_{x=0}^{\infty} r^x = \frac{1}{1-r}$$

$$\therefore E(X) = p q \frac{\partial}{\partial q} \frac{1}{1-q} = p q \frac{1}{(1-q)^2} = p q \frac{1}{p^2} = \frac{q}{p}$$

$$2) \text{ var}(X) = E(X^2) - [E(X)]^2 = \frac{q}{p^2}$$

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

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

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

2) The moment generating function (m.g.f.) of $X \sim \text{Geo}(p)$

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

$$\because \sum_{x=0}^{\infty} a^x = \left(\frac{1}{1-a} \right) \text{ by law } \Rightarrow \text{where; } a = q e^t$$

$$\therefore M_X(t) = p \left(\frac{1}{1 - q e^t} \right) \Rightarrow \therefore M_X(t) = \left(\frac{p}{1 - q e^t} \right) \quad , \text{ for } q e^t < 1$$

$$M'_X(t) = M'_X(t)|_{t=0}$$

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

$$\begin{aligned} M'_X(t) &= -p (1 - q e^t)^{-2} (-q e^t)|_{t=0} = p (1 - q e^0)^{-2} (q e^0) \\ &= \frac{pq}{(1-q)^2} = \frac{p}{q} \end{aligned}$$

$$M''_X(t) = p (1 - q e^t)^{-2} (q e^t) - 2p q e^t (1 - q e^t)^{-3} (-q e^t)|_{t=0}$$

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

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

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

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

3) The cumulative dist. function (c.d.f.) of $X \sim \text{Geo}(p)$.

$$F(k) = P(X \leq k) = \sum_{k=0}^x p q^k = p \sum_{k=0}^x q^k = p \left(\frac{1 - q^{x+1}}{1 - q} \right) = 1 - q^{x+1}$$

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ 1 - q^{x+1} & , \quad 0 \leq x < \infty \\ 1 & , \quad x \rightarrow \infty \end{cases} \quad \text{or; } F(x) = \begin{cases} 0 & x < 1 \\ 1 - q^x & 1 \leq x < \infty \\ 1 & x \rightarrow \infty \end{cases}$$

4) Mode of Geometric Distribution $X \sim \text{Geo}(p)$.

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

$$f(0; p) = p$$

$$f(1; p) = p q^1$$

$$f(2; p) = p q^2$$

$$\dots \quad , \quad \rightarrow \quad p > pq > pq^2 > \dots \quad , \therefore \text{mode} = 0$$

If one of the two numbers is less than one, then when multiplied together, the result is less than the value of one of them, such as p.

5) Memoryless (No memory) Property of Geometric Distribution

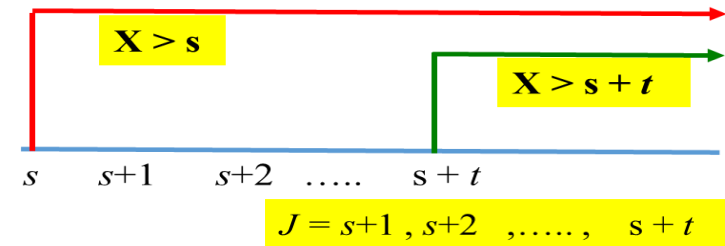
Theorem: Suppose that $X \sim \text{Geo}(p)$. Then for positive integers (s) , (t) , we have;

$$p(X > s + t \mid X > s) = p(X > t)$$

If an event has not occurred by time (s) , the probability that it will occur after an additional (t) time units is the same as the (unconditional) probability that it will occur after time (t) --- it forgot that it made it past time s !. (Reference: Mood and Graybill)

Proof:

$$\begin{aligned} p(X > s + t \mid X > s) &= \frac{p(X > s + t \cap X > s)}{p(X > s)} \\ &= \frac{p(X > s + t)}{p(X > s)}, \quad (t \text{ positive}) \\ &= \frac{1 - F(s + t)}{1 - F(s)} \\ &= \frac{1 - (1 - q^{s + t + 1})}{1 - (1 - q^{s + 1})} = q^t, \text{ free of } (s) \end{aligned}$$



$$\begin{aligned} \text{Or; } p(X > s + t \mid X > s) &= \frac{p(X > s + t)}{p(X > s)} = \frac{\sum_{x=s+t+1}^{\infty} p q^x}{\sum_{x=s+1}^{\infty} p q^x}, \rightarrow \sum_{x=n}^{\infty} r^x = \frac{r^n}{1-r}, \quad 0 < r < 1 \\ &= \frac{(q^{s+t+1})/(1-q)}{(q^{s+1})/(1-q)} = q^t \end{aligned}$$

$$\text{If, } p(X \geq s + t \mid X \geq s) = \frac{p(X \geq s + t)}{p(X \geq s)} = \frac{\sum_{x=s+t}^{\infty} p q^x}{\sum_{x=s}^{\infty} p q^x} = \frac{(q^{s+t})/(1-q)}{(q^s)/(1-q)} = q^t = p(X \geq t)$$

$$\text{Because, } p(X \geq t) = \sum_{x=t}^{\infty} p q^x = p \frac{q^t}{1-q} = q^t$$

Maximum Likelihood Estimation for Parameter of Geometric Distribution

Q// Find the value of p , which maximized p.d.f., or (for what value of p , the p.d.f. is maximum)
In a rssi from Geometric distⁿ Geo(p), with p.d.f ; $f(x;p) = p(1 - p)^{x-1}$, $x = 1,2,\dots$, find the m.l.e for p :

Sol: $X \sim \text{Geo}(p)$

First Case:

$$f(x;p) = p(1 - p)^{x-1} \quad , \quad x = 1,2, \dots$$

$\therefore X$'s are indep.

$$\begin{aligned} L(p) &= f(x_1, x_1, \dots, x_1; p) = \prod f(x_i; p) \\ &= p^n (1 - p)^{\sum x_i - n} \end{aligned}$$

$$\ln L(p) = n \ln(p) + (\sum x_i - n) \ln(1 - p)$$

$$\frac{\partial \ln L(p)}{\partial p} = \frac{n}{p} - \frac{(\sum x_i - n)}{(1 - p)} \quad , \quad \frac{\partial \ln L(p)}{\partial p} = 0$$

$$\frac{n}{p} - \frac{(\sum x_i - n)}{(1 - p)} = 0$$

$$\frac{n(1 - \hat{p}) - \hat{p}(\sum x_i - n)}{\hat{p}(1 - \hat{p})} = 0$$

$$n - n\hat{p} - \hat{p}\sum x_i + n\hat{p} = 0$$

$$n - \hat{p}\sum x_i = 0$$

$$\hat{p}\sum x_i = n \quad \hat{p}_{m.l.e} = \frac{n}{\sum x_i} = \frac{1}{\bar{X}}$$

$$\frac{\partial^2 \ln L(p)}{\partial p^2} = -\frac{n}{p^2} - \frac{(\sum x_i - n)}{(1 - p)^2} < 0$$

$\therefore \hat{p} = 1/\bar{X}$ is m.l.e for p .

Moment Method for Parameter of Geometric Distribution

$$X \sim \text{Goe}(p) \quad , \text{first case: } f(x; p) = p q^x \quad E(X) = \frac{1-p}{p} \quad ,$$

$$m_k = \frac{\sum x_i^k}{n} \quad , \quad M_k = E(X^k) \quad , \rightarrow \quad m_k = M_k$$

$$m_1 = \frac{\sum x_i^1}{n} = \bar{X} \quad , \quad M_1 = E(X) = \frac{1-p}{p}$$

$$m_1 = M_1$$

$$\bar{X} = \frac{1-p}{p} \quad \rightarrow \quad \bar{X} p = 1 - p \quad \rightarrow \quad \hat{p}_{\text{moment}} = \frac{1}{1 + \bar{X}}$$

$$X \sim \text{Goe}(p) \quad , \text{second case: } f(x; p) = p q^{x-1} \quad E(X) = \frac{1}{p} \quad ,$$

$$m_k = \frac{\sum x_i^k}{n} \quad , \quad M_k = E(X^k) \quad , \rightarrow \quad m_k = M_k$$

$$m_1 = \frac{\sum x_i^1}{n} = \bar{X} \quad , \quad M_1 = E(X) = \frac{1}{p}$$

$$m_1 = M_1$$

$$\bar{X} = \frac{1}{p} \quad \rightarrow \quad \bar{X} p = 1 \quad \rightarrow \quad \hat{p}_{\text{moment}} = \frac{1}{\bar{X}}$$

Note: In all distributions, it is not a requirement: Moment Method equal to the m.l.e Method.

For example, in Beta distribution, or in some cases we cannot use m.l.e. method.

Q// Is m.l.e $\hat{p} = \frac{1}{\bar{X}}$ unbiased estimator for p in *Geo.* distribution? $\hat{p} = \frac{n}{\sum x_i}$

H.W: Is m.l.e $\hat{p} = \frac{1}{1+\bar{X}} = \frac{n}{n+\sum x_i}$ unbiased estimator for p in *Geo.* distribution? $f(x; p) = p q^x$.

Exponential Family

Q// Is Geometric distribution $\text{Geo}(p)$ belongs to exponential family?

1) $f(x; p) = p q^x$, $x = 0, 1, \dots$

$$f(x; p) = \exp(\ln f(x; p))$$

$$= \exp[\ln p + x \ln(1 - p)]$$

$$A(\theta) = \ln(1 - p) \text{ , } B(x) = x \text{ , } C(\theta) = \ln p \text{ , } D(x) = 0$$

$\therefore f(x; p)$ of Geometric distribution belongs to exponential family.

In arssn;

$$f(x_1, x_2, \dots, x_n; p) = \exp(n \ln p + \ln(1 - p) \sum x_i)$$

$\therefore \sum B(X_i) = \sum x_i$ is suff. est. for p .

2) $f(x; p) = p q^{x-1}$, $x = 1, 2, \dots$

$$f(x; p) = \exp(\ln f(x; p))$$

$$= \exp[\ln p + (x - 1) \ln(1 - p)]$$

$$= \exp \left[\ln \left(\frac{p}{1-p} \right) + x \ln(1 - p) \right]$$

$$A(p) = \ln(1 - p) \text{ , } B(x) = x \text{ , } C(\theta) = \ln \left(\frac{p}{1-p} \right) \text{ , } D(x) = 0$$

$\therefore f(x; p)$ of Geometric distribution belongs to exponential family.

In arssn;

$$f(x_1, x_2, \dots, x_n; p) = \exp \left(n \ln \left(\frac{p}{1-p} \right) + \ln(1 - p) \sum x_i \right)$$

$\therefore \sum B(X_i) = \sum x_i$ is suff. est. for p

5) Negative Binomial Distribution

Independent Bernoulli trials are performed until (r) successes appear, define the r.v. X is the number of failure trials before getting the r -th success trial, then a r.v. X defined to have (N.B.) dist. if the p.d.f. of X given by:

$$p(x) = \begin{cases} C_x^{x+r-1} p^r q^x & , \quad x = 0, 1, 2, \dots \\ 0 & \text{o.w} \end{cases}$$

Where the parameters (r) and (p) satisfy [$r = 1, 2, \dots, 0 < p < 1$].

x : No. of failure trials before getting the r -th success.

r : No. of successes cases (fixed number).

Clearance: Let a coin tossed nine times, in ninth toss we get success trials, success trials (get 3 Heads). What is the probability of the following result?

$$\begin{aligned}
 p(TTT \underset{r=1}{H} T \underset{r=2}{H} TT \underset{r=3}{H}) &= \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \\
 &= C_x^{x+r-1} = C_6^8 \quad p^r = p^3 \quad q^x = q^6
 \end{aligned}$$

C_6^8 : No. of failure trials before getting the (3) success trials (H)