Distribution Theory

Department of Statistics & Informative

Ph.D. Students

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, …, b}.

Assumptions:

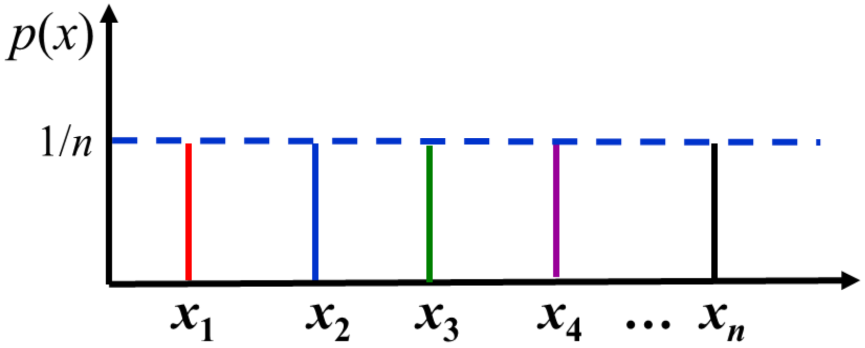
1) $X \in \{a, a+1, \ldots, b\}$

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 ~ (a, b) Probability Mass Function:

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

$$Let; \quad n = b - a + 1$$

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



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:

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

Where the parameter (n) is positive integer: X ~ 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;

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

$$F(x) = \begin{cases} 0 & , \ x < 1 \\ 1/n & , \ 1 \le x < 2 \\ 2/n & , \ 2 \le x < 3 \\ \cdot & \\ \cdot & \\ \dfrac{n-1}{n} & , \ n-1 \le x < n \\ 1 & , \ x \ge n \end{cases}$$

2) The mean and the variance of uniform distribution

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

$$Variance=v(X)=\ (n+1)\frac{b\ -\ \ a}{12}=\frac{(n+1)(n-1)}{12}$$

$$=\frac{n^2-1}{12}=\frac{(b-a+1)^2-1}{12}$$

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

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

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

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

4) Mode of uniform distribution

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

$\because$ 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:

$$\left.\begin{aligned} &1- \;\; p(X < m) = \sum_{x<m} p(x) \le \frac{1}{2} \\ &2- \;\; p(X \le m) = \sum_{x\le m} p(x) \ge \frac{1}{2} \end{aligned}\right\}$$

median = mean in D.U. dist.

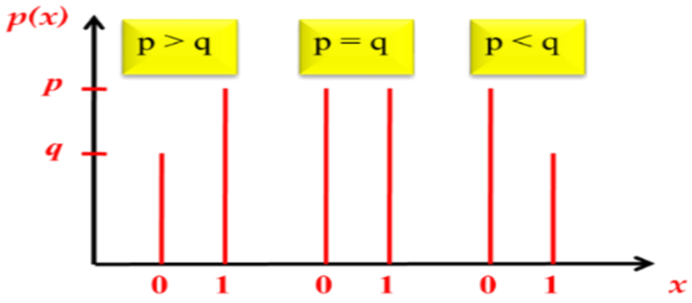
1. Bernoulli Distribution

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-defective ….etc. , with;

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

The p.m.f. of X;

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



The p.m.f. of Ber. Distribution

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 ~ Ber (1, p),

$$X \sim Ber\,(\underset{\text{No.of success trials}}{1}, \underset{\text{probability of success}}{p})$$

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

Properties of Bernoulli Distribution

a) The mean and the variance of X

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

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

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

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

b) The cumulative dist. function (c.d.f.) of X ~ Ber(1, *p*)

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

c) The moment generating function (m.g.f.) when X ~ Ber(1, *p*)

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

$$= \sum_{x=0,1} (pe^t)^x (q)^{1-x} = (pe^t)^0 (q)^{1-0} + (pe^t)^1 (q)^{1-1} = q + pe^t$$

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

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

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

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

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 ~ 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:

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

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

e) The Median of X ~ Ber(1, p) (Me)

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

∵ ½ is not within range, ∴ Bernoulli distribution does not have median.

Related Distributions

1) Binomial Dist. 2) Geometric Dist. 3) Negative Binomial Dist. 4) Beta Dist.

5) Catagorical Dist.

Estimation Theory for Bernoulli Distribution

Methods of Estimation

1. Moments Estimation Method

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

$$m_k = M_k$$

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

$M_1 = m_1 \rightarrow \therefore \; \hat{p} = \bar{X} \rightarrow$ 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 ~ Ber(p),

Let $X_1, X_2, \ldots, X_n$ denote a random sample from Bernoulli dist^n Ber(p), then the m.l.e for p.

$\because$ X ~ Ber(p)

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

$$\because X's \ are \ indep.$$

$$L(p) = f(x_1, x_1, \ldots, x_1;\ p) = \Pi f(x_i;p)$$

$$= p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$$

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

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

$$\frac{\Sigma x_i}{p\hat{}} - \frac{n-\Sigma x_i}{1-p\hat{}} = 0$$

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

$$\Sigma x_i - p\hat{}\Sigma x_i - np\hat{} + p\hat{}\Sigma x_{i_i} = 0$$

$$\Sigma x_i - np\hat{} = 0$$

$$\Sigma x_i = np\hat{} \qquad p\hat{}_{m.l.e} = \frac{\Sigma X_i}{n} = \overline{X}$$

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

$\therefore p\hat{} = \overline{X}$ is m.l.e for p → proportion of success (which gives the greatest value .

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

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

Properties of Estimator

1. Unbiasedness

If E(p^) = p , then an estimator is unbiased estimator for p.

$$E(\bar{X}) = E\left(\frac{\sum x_i}{n}\right) = \frac{n}{n} E(X) = p \quad , X \sim Ber(p) \; , E(X) = p$$

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

2. Unbiased in Limit

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

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

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;

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

or;

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

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

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. $\lim\limits_{n\to\infty} E(\hat{\theta})=\theta$ $\rightarrow$ unbiased in limit ,
2. $\lim\limits_{n\to\infty} v(\hat{\theta})=0$ $\rightarrow$ There is no error

$\lim\limits_{n\to\infty} E(\hat{p})=p$ 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:

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

$$= E\left(\hat{\theta} - \theta \mp E(\hat{\theta})\right)^2$$

$$= E\left(\{\hat{\theta} - E(\hat{\theta})\} + \{E(\hat{\theta}) - \theta\}\right)^2$$

$$= E\left(\hat{\theta} - E(\hat{\theta})\right)^2 + E\left(E(\hat{\theta}) - \theta\right)^2 + 2\, E\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)$$

$$= Var(\hat{\theta}) + b^2(\hat{\theta}) + zero$$

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

The Score Function

The score function is the partial derivative of *Ln* the function *f*(*x*;*θ*) 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; *θ*)) = zero (under Regularity Condition)

Proof:

$$E\big(S(X;\theta)\big) = \int_{R_X} s(x;\theta)\ f(x;\theta)\ dx = \int_{R_X} \frac{1}{f(x;\theta)} \frac{\partial}{\partial \theta} f(x;\theta)\, f(x;\theta)\, dx$$

$$= \int_{R_X} \frac{\partial}{\partial \theta} f(x;\theta)\ dx = \frac{\partial}{\partial \theta} \int_{R_X} f(x;\theta)\ dx = \frac{\partial}{\partial \theta}(1) = zero$$

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;

$$V(S(x;\theta) = E\Big(S(x;\ \theta) - E\big(S(x;\ \theta)\big)\Big)^2$$

$$= E\big(S(x;\ \theta)\big)^2 - \big(E(S(x;\ \theta)\big)^2 \quad , E(S) = 0$$

$$= E\big(S(x;\ \theta)\big)^2 \quad \rightarrow \text{Information}$$

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

If Fisher Information multiply by (*n*), we get; *n* I(*θ*) → F.I in a rss(*n*)

The Rao- Cramer Inequality

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

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

Notes:

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

a) $F.I = V(S) = -n\,E\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\,\theta^2}\right)$

b) $F.I = V(S) = n\,E\left(\frac{\partial\,Ln\,f(x;\theta)}{\partial\theta}\right)^2$

} $f(x;\,\theta)$ for single

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

d) $F.I = V(S) = \,E\left(\frac{\partial\,Ln\,L(\theta)}{\partial\theta}\right)^2$

} $L(x;\,\theta)$

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

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

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

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

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

4) $\because$ *X's are independent*

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

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?

Def^n: 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;

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

$$if \;\; eff = 1 \quad \Rightarrow T \;\; is \;\; called \;\; efficient \;\; estimator \;\; for \;\; \theta.$$

Def^n: 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).

$$V(T) = RCLB$$

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

for Bernoulli distribution X~ Ber(p)

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

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

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

F.I. in a rssn = $n\,\mathrm{E}\left(\frac{\partial \ln f(x;p)}{\partial p}\right)^2$

$$= n\,\frac{E(X-p)^2}{p^2(1-p)^2}$$

$$= \frac{n\,p\,(1-p)}{p^2(1-p)^2} = \frac{n}{p\,(1-p)}$$

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

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

$\therefore \hat{p} = \frac{\sum X_i}{n}$ 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 $X_1, X_2, \ldots, X_n$ be a rss$n$ from the dist^n 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 $X_1, X_2, \ldots, X_n$ be a rss$n$ from Bernoulli dist^n Ber(θ). Show that $\hat{\theta} = \sum X_i$ is sufficient estimator for the parameter θ.

$$\because X \sim Ber(\theta)$$

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

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

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

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

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

$$= \frac{\theta}{\theta^2} + \frac{(1-\theta)}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{(1-\theta)} = \frac{1-\theta+\theta}{\theta(1-\theta)} = \frac{1}{\theta(1-\theta)}$$

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

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

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

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

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

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

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

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

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

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

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

Second Method (Conditional)

Let $X_1, X_2, \ldots, X_n$ be a rss$n$ from the dist^n 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 $(X_1, X_2, \ldots, X_n)$ given $\hat{\theta}$ does not contain the parameter θ:

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

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:

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

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 $\sum x_i$. $\Rightarrow \sum X_i = n\bar{X}$.

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

for Bernoulli distribution

$$\because X \sim Ber(\theta)$$

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

$$f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \quad \} \times \frac{C^{n}_{\sum x_i}}{C^{n}_{\sum x_i}}$$

$$= C^{n}_{\sum x_i}\, \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \times \frac{1}{C^{n}_{\sum x_i}} \quad , \; free \;\; of \;\; \theta$$

$$= g(\hat{\theta} = \sum x_i; \theta) \times H(x) \qquad \Rightarrow \therefore \hat{\theta} = \sum x_i \;\; is \;\; suff \;\; est \;\; for \;\; \theta$$

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:

$$f(x;\, \theta) = exp(\ln f(x;\, \theta))$$

$$= exp\big(A(\theta)\, B(x) + C(\theta) + S(x)\big) \qquad \dots.(1)$$

$$or; \qquad = g(\theta)\, h(x)\, e^{A(\theta)\, B(x)} \qquad \dots.(2)$$

Such that: A(θ) 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?

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

$$= \exp\left(x\ln(\theta) + (1-x)\ln(1-\theta)\right)$$

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

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

$\therefore$ 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;

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

where (n) and (p) are positive parameters, such that ($0 \le p \le 1$), n: No. of trials (positive integer), i.e., X ~ 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 ~ b(n, p), then; The mean and the variance of X.

$$mean = E(X) = \sum_{x=0}^{n} x\ p(x) = \sum_{x=0}^{n} x\, C_x^n\, p^x q^{n-x} = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

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

Let;

$$\left.\begin{matrix} y = x-1 \ , \ m = n-1 \\ x = y+1 \ , \ n = m+1 \end{matrix}\right\} \qquad Range: \begin{cases} x = 0 \Rightarrow y = -1 \ , \ x = n \Rightarrow y = n-1 \\ \therefore y = -1, 0, 1, 2, \ldots\ldots, n-1 \end{cases}$$

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

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

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

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

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

2) The moment generating function (m.g.f.) of X ~ bin(n, p)

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

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

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

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

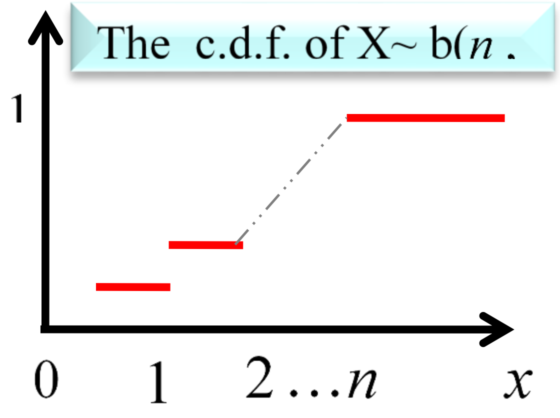
$$\therefore M_X(t) = (pe^t + q)^n$$

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 ~b(*n*, *p*).

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



4) Addition Property

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

Proof:

$$M_Y(t) = E\,e^{tY} = E\,e^{t\sum X_i} = E\,e^{\sum t X_i} = E(e^{tX_1}\,e^{tX_2}\,.......e^{tX_n})$$

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

For example; $Y = X_1 + X_2 + X_3$, where $X_1 \sim$ b($n1$,p), $X_2 \sim$ b($n2$,p), $X_3 \sim$ b($n3$,p)

$$\therefore Y = \sum_{i=1}^{n} X_i \sim b(\sum n_i, p)$$

→ $Y \sim b\,(n_1 + n_2 + n_3, p)$

5) The mode of Binomial Distribution

If the mode of distribution of X is unique?

Mode of X(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: ε = fraction $\longrightarrow$ then; $\boldsymbol{m = (n + 1)\, p + \varepsilon}$

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

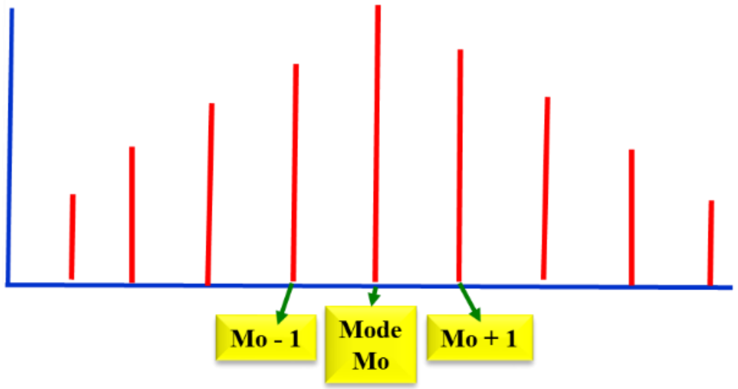
Mode integer = (m_1, m_2)

Mode fraction = (m)

Ex: $x = 6.1 \longrightarrow$ mode = 6

$x = 6.99 \longrightarrow$ mode = 6

if $x = 6$ (integer) $\longrightarrow$ mode = 6 , 5



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

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

$\therefore m = 4$

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:

Mean – Mode = 3(Mean – Median)

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

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

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

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

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

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:

Mean – Mode = 3(Mean – Median)

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

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

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

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

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 Bin(*n*, *p*) distribution $\rightarrow \lambda = np$ become Poisson distribution, Bin(*n*, *p*) ~ 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 $n + 1$:

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

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

Q// Is Binomial distribution belongs to exponential family?

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

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

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

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

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

∴ Binomial distribution belongs to exponential family.

Let $X_1, X_2, \ldots, X_n$ be a rss$n$ from Binomial dist^n Bin(m, θ), find the m.l.e for θ.

$$X \sim Bin(m,\theta)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Let $X_1, X_2, \ldots, X_n$ be a rss$n$ from Binomial dist[n] Bin(m, θ), 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.

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

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

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

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

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

Only $x = 0$
E(X) = E(0) = 0

Because the distribution of X is complete.

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

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

Ex: Let X be a random variable from Binomial distribution X ~Bin(n, θ). 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 for all $\theta \in \Omega$

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

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

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

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

$\therefore u(0) = u(1) = \cdots = u(n) = 0 \quad \rightarrow u(x) = 0 \ \forall x$

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

Multinomial Distribution

$(x+y)^n = Binomial$

$(x+y+z)^n = Trynomial$

$(x_1+x_2+\cdots+x_k)^n = Multinomial$

$$f(x_i;\ p_i) = \frac{n!}{x_1!x_2!\ldots.x_k!}\, p_1^{x_1} p_2^{x_2} \ldots.p_k^{x_k} \quad , \quad n = \sum_{i=1}^{k} x_i \ , \quad \sum_{i=1}^{k} p_i = 1$$

H.W: for each (p_i) find one (m.l.e).

$$p_j = 1 - \sum_{\substack{i\neq j \\ i=1}}^{k} p_i$$

Trinomial Distribution

When k = 3

$$f(x,y;\ p_i) = \frac{n!}{x!\ y!\,(n-x-y)!}\, p_1^x p_2^y (1-p_1-p_2)^{n-x-y} \quad ,$$

$$Or:\ \ f(x_1,x_2,x_3;\ p_i) = \frac{n!}{x_1!\,x_2!\ x_3!}\, p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

H.W: for each (p_i) find one (m.l.e).

$(p_1+p_2)^n = Binomial$

$(p_1+p_2+p_3)^n = Trynomial$

$(p_1+p_2+\cdots+p_k)^n = Multinomial$

H.W: Let $(X, Y) \sim$ Trinomial dist^n $\text{Tri}(n, \theta_1, \theta_2)$, find the m.l.e for θ_1, θ_2.

$$f(x, y;\ \theta_1, \theta_2) = \frac{n!}{x!\ y!\ (n-x-y)!}\ \theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y}$$

$$L(\theta_1, \theta_2) = \left[\prod_{j=1}^m \frac{n!}{x_j\, y_j\, (n-x_j-y_j)!}\right] \theta_1^{\sum_{j=1}^m x_j} \theta_2^{\sum_{j=1}^m y_j} (1-\theta_1-\theta_2)^{mn-\sum_{j=1}^m x_j - \sum_{j=1}^m y_j}$$

$$lnL(\theta_1, \theta_2) = \ln[k] + \sum_{j=1}^m x_j \ln\theta_1 + \sum_{j=1}^m y_j \ln\theta_2 + \left(mn - \sum_{j=1}^m x_j - \sum_{j=1}^m y_j\right)\ln(1-\theta_1-\theta_2)$$

$$\frac{\partial lnL(\theta_1, \theta_2)}{\partial\theta_1} = \frac{\sum_{j=1}^m x_j}{\theta_1} - \frac{\left(mn - \sum_{j=1}^m x_j - \sum_{j=1}^m y_j\right)}{(1-\theta_1-\theta_2)} = 0$$

$$\frac{\partial lnL(\theta_1, \theta_2)}{\partial\theta_2} = \frac{\sum_{j=1}^m y_j}{\theta_2} - \frac{\left(mn - \sum_{j=1}^m x_j - \sum_{j=1}^m y_j\right)}{(1-\theta_1-\theta_2)} = 0$$

$$\frac{\sum_{j=1}^m x_j}{\theta_1} = \frac{\left(mn - \sum_{j=1}^m x_j - \sum_{j=1}^m y_j\right)}{(1-\theta_1-\theta_2)} \quad \ldots.(1)$$

$$\frac{\sum_{j=1}^m y_j}{\theta_2} = \frac{\left(mn - \sum_{j=1}^m x_j - \sum_{j=1}^m y_j\right)}{(1-\theta_1-\theta_2)} \quad \ldots.(2)$$

$$\therefore\ \frac{\sum_{j=1}^m x_j}{\theta_1} = \frac{\sum_{j=1}^m y_j}{\theta_2} \ \to \theta_2 \sum_{j=1}^m x_j = \theta_1 \sum_{j=1}^m y_j \quad \ldots.(3)$$

from (1)

$$\sum_{j=1}^m x_j - \theta_1 \sum_{j=1}^m x_j - \theta_2 \sum_{j=1}^m x_j = mn\theta_1 - \theta_1 \sum_{j=1}^m x_j - \theta_1 \sum_{j=1}^m y_j$$

$$\therefore\ \theta_2 = \frac{\sum_{j=1}^m x_j - mn\theta_1 + \theta_1 \sum_{j=1}^m y_j}{\sum_{j=1}^m x_j} \quad , \qquad \textit{put it to } (3)$$

$$\frac{\sum_{j=1}^m x_j - mn\theta_1 + \theta_1 \sum_{j=1}^m y_j}{\sum_{j=1}^m x_j} \sum_{j=1}^m x_j = \theta_1 \sum_{j=1}^m y_j$$

$$\sum_{j=1}^m x_j - mn\theta_1 = 0 \ \to \hat{\theta}_1 = \frac{\sum_{j=1}^m x_j}{mn}$$

In the same way; $\hat{\theta}_2 = \frac{\sum_{j=1}^m y_j}{mn}$

Lagrange Function

Optimize; $\min Z = f(x_1, x_2, \dots, x_n)$

$G(x_1, x_2, \dots, x_n)$ $\quad a_1 x_{i1} + a_2 x_{i2} + \cdots + a_n x_{in} \lesseqgtr b \quad , i = 1,2,\dots,m$

$L(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n) + \lambda\,(G(x_1, x_2, \dots, x_n) - b)$

$$\frac{\partial L}{\partial x_j} = 0 \quad , j = 1,2,\dots,n$$

$$\frac{\partial L}{\partial \lambda_i} = 0 \quad , i = 1,2,\dots,m$$

H.W: Find the m.l.e for p_1, p_2 for the second form of Trinomial dist^n. by using Lagrange method.

$$f(x_1, x_2, x_3;\ p_i) = \frac{n!}{x_1!\,x_2!\ x_3!}\ p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$L(p_1, p_2) = \left[\prod_{j=1}^{m} \frac{n!}{x_j\ y_j\ z_j!}\right] p_1^{\sum_{j=1}^m x_j} p_2^{\sum_{j=1}^m y_j}\ p_3^{\sum_{j=1}^m z_j}$

$lnL(p_1, p_2) = \ln[s] + \sum_{j=1}^m x_j \ln(p_1) + \sum_{j=1}^m y_j \ln(p_2) + \sum_{j=1}^m z_j \ln(p_3) \qquad , \sum_{i=1}^3 p_i = 1$

To solve this we use Lagrange function;

$La(p) = \ln[s] + \sum_{j=1}^m x_j \ln(p_1) + \sum_{j=1}^m y_j \ln(p_2) + \sum_{j=1}^m z_j \ln(p_3) - \lambda(\sum_{i=1}^3 p_i - 1)$

$\frac{\partial La}{\partial p_1} = \frac{\sum_{j=1}^m x_j}{p_1} - \lambda = 0 \quad \rightarrow \frac{\sum_{j=1}^m x_j}{p_1} = \lambda \quad \rightarrow \sum_{j=1}^m x_j = \lambda\, p_1$

$\frac{\partial La}{\partial p_2} = \frac{\sum_{j=1}^m y_j}{p_2} - \lambda = 0 \quad \rightarrow \frac{\sum_{j=1}^m y_j}{p_2} = \lambda \quad \rightarrow \sum_{j=1}^m y_j = \lambda\, p_2$

$\frac{\partial La}{\partial p_3} = \frac{\sum_{j=1}^m z_j}{p_3} - \lambda = 0 \quad \rightarrow \frac{\sum_{j=1}^m z_j}{p_3} = \lambda \quad \rightarrow \sum_{j=1}^m z_j = \lambda\, p_3$ **+**

$\frac{\partial La}{\partial \lambda} = \sum_{i=1}^3 p_i - 1 = 0 \quad \rightarrow \sum_{i=1}^3 p_i = 1$ $\qquad \boldsymbol{n = \lambda}$

$\therefore\ \hat{p}_1 = \frac{\sum_{j=1}^m x_j}{\sum_{j=1}^m x_j + \sum_{j=1}^m y_j + \sum_{j=1}^m z_j} = \frac{\sum_{j=1}^m x_j}{n}\ , \hat{p}_2 = \frac{\sum_{j=1}^m y_j}{n}\ , \rightarrow \hat{p}_3 = \frac{\sum_{j=1}^m z_j}{n}$

H.W: Find the m.l.e for p_i for the Multinomial dist^n.

$$L(p_1, p_2, \dots, p_k) = \left[\prod_{j=1}^{m} \frac{n!}{x_{j1}!\, x_{j2}! \dots x_{jk}!}\right] p_1^{\sum_{j=1}^{m} x_{j1}} p_2^{\sum_{j=1}^{m} x_{j2}} \dots p_k^{\sum_{j=1}^{m} x_{jk}}$$

$$lnL = \ln[s] + \sum_{j=1}^{m} x_{j1} \ln(p_1) + \sum_{j=1}^{m} x_{j2} \ln(p_2) + \cdots + \sum_{j=1}^{m} x_{jn} \ln(p_k) \qquad , \sum_{i=1}^{k} p_i = 1$$

$$= \ln[s] + \sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}\, p_i$$

$$La(p_1, p_2, \dots, p_k, \lambda) = \ln[s] + \sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}\, p_i - \lambda\left(\sum_{i=1}^{k} p_i - 1\right)$$

$$\frac{\partial La}{\partial p_i} = \frac{\sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}}{p_i} - \lambda = 0 \quad \rightarrow \lambda\, p_i = \sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}$$

$$\frac{\partial La}{\partial \lambda} = \sum_{i=1}^{k} p_i - 1 = 0 \qquad \rightarrow \sum_{i=1}^{k} p_i = p_1 + p_2 + \cdots + p_k = 1$$

$$\lambda\, p_i = \sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}$$

$$\lambda\, p_1 = \sum_{j=1}^{m} x_{1j}$$

$$\lambda\, p_2 = \sum_{j=1}^{m} x_{2j}$$

$$\dots\dots$$

$$\lambda\, p_k = \sum_{j=1}^{m} x_{kj} \quad +$$

$$\boldsymbol{\lambda = \sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}}$$

$$\hat{p}_i = \frac{\sum_{j=1}^{m} x_{ij}}{\sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij}}$$

Covariance of Multinomial Distribution Cov(X_1, X_2)

$$E(X_i) = np_i$$

$$E(X_1X_2) = \sum_{x_k} \dots \sum_{x_2} \sum_{x_1} x_1 x_2 f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k)$$

$$= \sum_{x_k=0}^{n} \dots \sum_{x_2=0}^{n} \sum_{x_1=0}^{n} x_1 x_2 \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$= \sum_{x_k=0}^{n} \dots \sum_{x_3=0}^{n} \sum_{x_2}^{n} \sum_{x_1}^{n} \frac{n!}{(x_1-1)!(x_2-1)! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$= n(n-1)p_1p_2 \sum_{x_k}^{n} \dots \sum_{x_2=2}^{n} \sum_{x_1=2}^{n} \frac{(n-2)!}{(x_1-1)!(x_2-1)! \dots x_k!} p_1^{x_1-1} p_2^{x_2-1} \dots p_k^{x_k}$$

$let;$ $y_1 = x_1 - 1$, $y_2 = x_2 - 1$, $n! = n(n-1)(n-2)!$, $N! = (n-2)!$

$$E(X_1X_2) = n(n-1)p_1p_2 \sum_{y_k=0}^{N} \dots \sum_{y_2=0}^{N} \sum_{y_1=0}^{N} \frac{N!}{y_1! y_2! \dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}$$

$$\therefore\ E(X_1X_2) = n(n-1)p_1p_2$$

$$\therefore\ Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = n(n-1)p_1p_2 - (np_1)(np_2) = -np_1p_2$$

Moment Generation Function for Multinomial Distribution

[Ameer Hanna p306; Roussas p158]

$$M_X(t_1,t_2,\dots,t_k) = E\left(e^{\sum_{i=1}^n t_iX_i}\right) = \left[\sum_{i=1}^n p_ie^{t_i}\right]^n$$

$$M_X(t_1,t_2,\dots,t_k) = E\left(e^{\sum_{i=1}^n t_iX_i}\right) = \sum_x e^{\sum_{i=1}^n t_iX_i}.\frac{n!}{\prod_{i=1}^k x_i!}\prod_{i=1}^k p_i^{x_i}$$

$$= \sum_x \frac{n!}{\prod_{i=1}^k x_i!}.\prod_{i=1}^k (p_ie^{t_i})^{x_i}$$

According to the binomial theorem of degree (n) for any real numbers $a_1,a_2,\dots,a_k$

$$(a_1+a_2+\cdots+a_k)^n = \sum_x \frac{n!}{\prod_{i=1}^k x_i!}\prod_{i=1}^k a_i^{x_i} \quad, \; put \;\; a_i = p_ie^{t_i}$$

$$(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_ke^{t_k})^n = \sum_x \frac{n!}{\prod_{i=1}^k x_i!}\prod_{i=1}^k (p_ie^{t_i})^{x_i}$$

$$\therefore\; M_X(t_1,t_2,\dots,t_k) = \left[\sum_{i=1}^n p_ie^{t_i}\right]^n$$

Let; $K_x(t) = n\ln\sum_{i=1}^n p_ie^{t_i}$

$$Mean\;(X_i) = M_{X_i}(t_1,t_2,\dots,t_k) = \left.\frac{\partial K_x(t)}{\partial\, t_i}\right|_{t_1,t_2,\dots,t_k=0} = \left.\frac{np_ie^{t_i}}{\sum_{i=1}^n p_ie^{t_i}}\right|_{t_1,t_2,\dots,t_k=0} = np_i$$

$i = 1,2,..,k$

$$Var(X_i) = \left.\frac{\partial^2 K_x(t)}{\partial\, t_i^2}\right|_{t_1,t_2,\dots,t_k=0} = np_i\left[\frac{(\sum_{i=1}^n p_ie^{t_i})e^{t_i}-(e^{t_i})(p_ie^{t_i})}{(\sum_{i=1}^n p_ie^{t_i})^2}\right]_{t_1,t_2,\dots,t_k=0} = np_i(1-p_i)$$

$$Cov\left(X_i,X_j\right) = \sigma_{ij} = \left.\frac{\partial^2 K_x(t)}{\partial\, t_i\;\partial\, t_j}\right|_{t_1,t_2,\dots,t_k=0} = np_ie^{t_i}\left[-\left(\sum_{i=1}^n p_ie^{t_i}\right)^{-2}p_je^{t_j}\right]\Big|_{t_1,t_2,\dots,t_k=0}$$

$$\therefore\; \sigma_{ij} = -np_ip_j \quad , \quad i,j = 1,2,3,\dots,n$$

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 $\{0,1,2,3,\ldots\}$or$\{1,2,3,\ldots\}$ 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:

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

The second form:

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

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

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 ~ Geo(*p*).

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

$$= pq \sum_{x=0}^{\infty} \frac{\partial}{\partial q} q^{x} \qquad , because \ \frac{\partial}{\partial q} q^{x} = x\, q^{x-1}$$

$$= pq \frac{\partial}{\partial q} \sum_{x=0}^{\infty} q^{x} \qquad , by\ low \sum_{x=0}^{\infty} r^{x} = \frac{1}{1-r}$$

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

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

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

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

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

2) The moment generating function (m.g.f.) of X ~ Geo(p)

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

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

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

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

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

$$M'_X(t) = -p\,(1-qe^t)^{-2}\,(-qe^t)|_{t=0} = p\,(1-qe^0)^{-2}\,(qe^0)$$

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

$$M''_X(t) = p\,(1-qe^t)^{-2}\,(qe^t) - 2p\,qe^t(1-qe^t)^{-3}\,(-qe^t)|_{t=0}$$

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

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

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

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

3) The cumulative dist. function (c.d.f.) of X ~Geo(*p*).

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

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

4) Mode of Geometric Distribution X ~Geo(*p*).

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

$f(0;p) = p$

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

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

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.

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

5) Memoryless (No memory) Property of Geometric Distribution

Theorem: Suppose that X ~ 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)

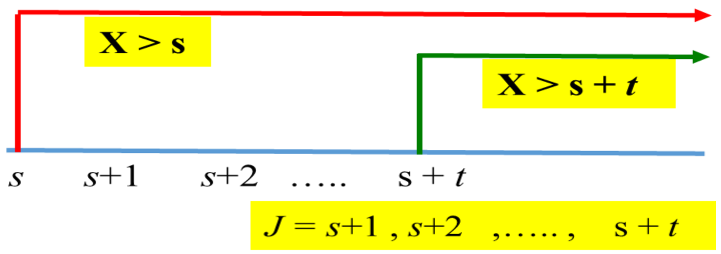
Proof:

$$p(X > s + t \mid X > s) = \frac{p(X>s+t \cap X>s)}{p(X>s)}$$

$$= \frac{p(X>s+t)}{p(X>s)} \quad , (t \text{ positive})$$

$$= \frac{1-F(s+t)}{1-F(s)}$$

$$= \frac{1-(1-q^{s+t+1})}{1-(1-q^{s+1})} = q^t \quad , \text{free of (s)}$$



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

$$= \frac{(q^{s+t+1})/(1-q)}{(q^{s+1})/(1-q)} = q^t$$

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

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 rss$n$ from Geometric dist^n Geo(p), with p.d.f ; $f(x;p) = p(1-p)^{x-1}$, $x = 1,2,\ldots\ldots$, find the m.l.e for p:

Sol: X ~ Geo(p)

First Case:

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

$$\because X's \ are\ indep.$$

$$L(p) = f(x_1, x_1, \ldots, x_1;\ p) = \Pi\, f(x_i;p)$$

$$= p^n (1-p)^{\Sigma x_i - n}$$

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

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

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

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

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

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

$$\hat{p}\Sigma x_i = n \qquad \hat{p}_{m.l.e} = \frac{n}{\Sigma X_i} = \frac{1}{\overline{X}}$$

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

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