BINOMIAL DISTRIBUTION

Presented
By
Pradip Panda
Asstt. Professor, Department of Statistics
Serampore College

CONTENTS

Trial: The word ‘trial ’ means an attempt to produce a particular event which is neither certain nor impossible.

For example: Throwing a die, tossing a coin, etc.

Independent trials:

Trials are said to be independent in relation to an event if the probability of the event in any trial is not affected by the supplementary information about the results of any number of other trials.

Example of independent trial: Tossing a coin repeatedly and noting if head appears or not.

Example of dependent trial: Drawing balls from a bag, containing a specified number of red and blue balls, one by one without replacement and noting their colour.

BERNOULLIAN TRIALS

A series of trials are said to be Bernoullian series of trials if

- the possible results of any trial can be divided into two classes i.e., 'success' and 'failure '.
- the trials are independent.
- the probability of success (or that of failure) remains constant from trial to trial.

n tosses of an unbiased coin will give a set of Bernoullian trials, where getting head may be regarded as a success with probability ½ in each trial.

BERNOULLI DISTRIBUTION

X (a random variable)

1 (or success) with probability p.

0 (or failure) with probability q = 1-p.

Probability – mass function of X is

$$P(X = x) = p^{x}(1-p)^{1-x};\ x = 0,1;\ 0 < p < 1$$
$$= 0 \qquad , \quad elsewhere.$$

$$E(X) = \sum_{x=0}^{1} x\, P(X = x) \;= \sum_{x=0}^{1} x.p^{x}(1-p)^{1-x}$$
$$= 0.(1-p) + 1.p = p$$

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

$$So, Var(X)=E(X^2)-E^2(X) = p - p^2 = p(1-p) = pq.$$

If X is the number of successes in n Bernoullian trials, then X ~ Bin(n,p).

$$X = X_1 + X_2 + \dots\dots\dots + X_i + \dots\dots\dots + X_n$$

Number of successes in n trials.

Bernoulli variable, taking value 0 or 1, with prob. p of success and q of failure.

$$E(X)=E(X_1)+E(X_2)+\dots\dots\dots+E(X_i)+\dots\dots\dots+E(X_n) = np.$$

$$Var(X)=\sum_{i=1}^{n}V(X_i) = n.p.q$$

BINOMIAL DISTRIBUTION

If X ~ Bin(n,p), then the probability – mass function is given by

$$P(X = x) = n_{c_x} p^x q^{n-x}, x = 0,1,2,\ldots\ldots, n; 0 < p < 1;$$
$$= \quad 0 \qquad , \quad elsewhere. \qquad p + q = 1$$

Properties: $i)\ Mean(\mu) = np$

$ii)\ Var(X) = \sigma^2 = npq; \therefore s.d.(\sigma) = \sqrt{npq}; q = 1 - p.$

$iii)\ \mu_3 = npq(q-p); \mu_4 = 3n^2p^2q^2 + npq(1-6pq)$

$iv)$ *The distribution is bimodal when* $(n+1)p$ *is an integer, the two modes being* $(n+1)p$ *and* $(n+1)p-1$; *again the distribution is unimodal when* $(n+1)p$ *is fractional, the mode being the greatest integer contained in* $(n+1)p$.

$$v)\ \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq}; \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq}.$$

vi) A measure of skewness is $\gamma_1 = \sqrt{\beta_1} = \dfrac{q-p}{\sqrt{npq}}$.

Distribution

- Positively skewed, when $p<q$ or $p<1/2$.
- Symmetric, when $p=q$ or $p=1/2$.
- Negatively skewed, when $p>q$ or $p>1/2$.

vii) A measure of kurtos is $\gamma_2 = \beta_2 - 3 = \dfrac{1-6pq}{npq}$.

Distribution:

- Leptokurtic, when $1\text{-}6pq>0$ or $pq<1/6$.
- Mesokurtic, when $1\text{-}6pq=0$ or $pq=1/6$.
- Platykurtic, when $1\text{-}6pq<0$ or $pq>1/6$

RECURSION RELATION FOR MOMENTS

Statement:

$$\mu_{r+1} = pq\{nr\mu_{r-1} + \frac{d\mu_r}{dp}\}$$

Proof:

$$\mu = np.$$

$$\mu_r = E(X - np)^r = \sum_{x=0}^{n} (x - np)^r n_{c_x} p^x q^{n-x}$$

$$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} n_{c_x} [r(x - np)^{r-1}(-n) p^x q^{n-x} +$$

$$(x - np)^r \{xp^{x-1} q^{n-x} - (n - x) p^x q^{n-x-1}\}]$$

$$= -nr\sum_{x=0}^{n} (x - np)^{r-1} p^x q^{n-x} + \sum_{x=0}^{n} (x - np)^r p^x q^{n-x} \{\frac{x}{p} - \frac{n-x}{q}\}$$

$$= -nr\mu_{r-1} + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^r n_{c_x} p^x q^{n-x}\{xq-(n-x)p\}$$

$$= -nr\mu_{r-1} + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^r n_{c_x} p^x q^{n-x}\{x(p+q)-np\}$$

$$= -nr\mu_{r-1} + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^{r+1} n_{c_x} p^x q^{n-x}$$

$$= -nr\mu_{r-1} + \frac{\mu_{r+1}}{pq}$$

$$\therefore \quad \mu_{r+1} = pq\{nr\mu_{r-1} + \frac{d\mu_r}{dp}\}$$

$$Now,\ \mu_1' = np.$$

$$\mu_2 = pq\{n.\mu_0 + \frac{d\mu_1}{dp}\} = npq, \because \mu_0 = 1\ and\ \mu_1 = 0.$$

$$\mu_3 = pq\{2n.\mu_1 + \frac{d\mu_2}{dp}\} = pq\{\frac{dnp(1-p)}{dp}\}.$$

$$= pq\{n - 2np\} = pq\{n(p+q) - 2np\}$$

$$= pq\{n - 2np\} = pq\{n(p+q) - 2np\}$$

$$= pq\{nq - np\} = npq\{q - p\}$$

MEAN DEVIATION ABOUT MEAN

$$MD_{\mu} = E\{|X - \mu|\} = \sum_{x=0}^{n} |x - np| . n_{c_x} p^x (1 - p)^{n-x}$$

$$= 2 \sum_{x=k+1}^{n} (x - np) n_{c_x} p^x (1 - p)^{n-x}, \; where \; k = [np]$$

$$= 2 \sum_{x=k+1}^{n} \{x(p+q) - np\} n_{c_x} p^x (q)^{n-x}$$

$$= 2 \sum_{x=k+1}^{n} \{xq - (n-x)p\} \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= 2 \sum_{x=k+1}^{n} \left\{ \frac{n! \, p^x q^{n-x+1}}{(x-1)!(n-x)!} - \frac{n! \, p^{x+1} q^{n-x}}{x!(n-x-1)!} \right\}$$

$$= 2\sum_{x=k+1}^{n}\left(v_x - v_{x+1}\right), \textit{ where } v_x = \frac{n!\,p^x q^{n-x+1}}{(x-1)!(n-x)!}$$

$$= 2v_{k+1}, \left(\because v_{n+1} = 0\right)$$

$$= 2.\frac{n!\,p^{k+1}q^{n-k}}{k!(n-k-1)!} = 2npq.\frac{(n-1)!\,p^k q^{n-1-k}}{k!(n-1-k)!}$$

$$= 2npq.(n-1)_{c_k}\, p^k q^{n-1-k}$$

MODE OF BINOMIAL DISTRIBUTION

$$Since,\ X \sim Bin(n, p),$$

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

$$= 0 \quad , otherwise \qquad 0 < p < 1;$$

$$q = 1 - p$$

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

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

$$= \frac{n-x+1}{x} \cdot \frac{p}{q} \quad = \frac{np - xp + p + xq - xq}{xq}$$

$$= 1 + \frac{(n+1)p - x}{xq}, \quad x = 1,2,3,\ldots\ldots\ldots, n$$

$Case-1:$ $Let\ (n+1)p\ be\ an\ integer\ and\ is\ equal\ to\ m.$

$$So,\ \frac{f(x)}{f(x-1)} = 1 + \frac{m-x}{xq},\ x = 1,2,3,..........,n$$

$$\frac{f(x)}{f(x-1)} > 1,\ when\ x = 1,2,............,(m-1)$$

$$\therefore f(0) < f(1) < .... < f(m-1)$$

$$\frac{f(x)}{f(x-1)} = 1,\ when\ x = m.$$

$$Then\ f(m-1) = f(m).$$

$$\frac{f(x)}{f(x-1)} < 1,\ when\ x = (m+1), (m+2), ..., n$$

$$\therefore\ f(m) > f(m+1) > .... > f(n)$$

$Combining\ above\ three\ results\ it\ is\ found\ that$

$$f(0) < f(1) < ...... < f(m-1) = f(m) >$$
$$f(m+1) > .... > f(n)$$

The distribution is bimodal and modes are $(m-1)$ and m.

Case – 2:

Let $(n+1)p$ be not an integer and is equal to $m'+f$. Where, m' is the integral part and f is the fractional part.

$$So,\ \frac{f(x)}{f(x-1)} = 1 + \frac{m'+f-x}{xq},\ x = 1,2,3,\ldots\ldots,n$$

$$\frac{f(x)}{f(x-1)} > 1,\ when\ x = 1,2,\ldots\ldots\ldots\ldots, m'$$

$$\therefore\ f(0) < f(1) < \ldots\ldots\ldots\ldots < f(m')$$

$$\frac{f(x)}{f(x-1)} \neq 1,\ \text{since}\ x\ takes\ \text{integral}\ values\ only.$$

$$\frac{f(x)}{f(x-1)} < 1,\ when\ x = (m'+1),(m'+2),\ldots\ldots, n$$

$$\therefore\ f(m') > f(m'+1) > \ldots\ldots\ldots\ldots\ldots\ldots > f(n)$$

Combining above three results it is found that

$$f(0) < f(1) < .... < f(m') > f(m'+1) > .... > f(n)$$

The distribution is unimodal and mode is m'.

FITTING OF BINOMIAL DISTRIBUTION

- A survey of 320 families with 5 children each revealed the following distribution:

No. of Boys	0	1	2	3	4	5
No. of families	12	40	88	110	56	14

Fit a binomial distribution to the data and find the expected frequencies for different classes.

Solution: Let X denote the no. of boys in a family of 5 children, p be the probability that a child is a boy.

X ~ Bin (5,p) . Here, p is unknown. p is estimated by the method of moments.

$$\therefore \bar{x} = \frac{\sum xf}{\sum f} = \frac{840}{320} = 2.625$$

$$So, by\ the\ method\ of\ moments\ n\hat{p} = \bar{x}\ or,\ \hat{p} = \frac{\bar{x}}{n}.$$

$$\hat{p} = \frac{2.625}{5} = 0.525\ and\ \hat{q} = 1 - \hat{p} = 0.475.$$

$$Now,\ f(x) = 5_{c_x} (0.525)^x (0.475)^{5-x},\ x = 0(1)5.$$

Expected frequency of x is 320 x f(x).

Recursion relation is

$$f(x) = \frac{5-x+1}{x} \cdot \frac{\hat{p}}{\hat{q}} \cdot f(x-1) = \frac{6-x}{x} \cdot \frac{\hat{p}}{\hat{q}} \cdot f(x-1),\ x > 0.$$

$$f(0) = (\hat{q})^5 = (0.475)^5 = 0.02418 \quad and\ \frac{\hat{p}}{\hat{q}} = 1.10526.$$

X (1)	(2)	(3)	f(x) = f(x-1)xcol.(3) (4)	Expected frequency =320xf(x) (5)
0	-	-	0.02418	7.74
1	5	5.52630	0.13362	42.76
2	2	2.21052	0.29537	94.52
3	1	1.10526	0.32646	104.46
4	0.5	0.55263	0.18041	57.73
5	-	-	*0.03996	12.79
Total			1	320

* Obtained by P(X = 5) = 1-[P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)]

Problem:

$$If\ \frac{f(x)}{f(x-1)} = \frac{n-x+1}{x}.\frac{p}{q}\ , x = 1,2,3,......n;$$

$$then\ \ find\ \ the\ \ p.m.f\,.$$

Solution:

$$f(1) = n.\frac{p}{q}.f(0) = n_{C_1}.\frac{p}{q}.f(0)$$

$$f(2) = \frac{n-1}{2}.\frac{p}{q}.f(1) = n_{C_2}.(\frac{p}{q})^2.f(0)$$

$$f(3)=\frac{n-2}{3}.\frac{p}{q}.f(2)=n_{C_3}.(\frac{p}{q})^3.f(0)$$

....

....

$$f(n)=n_{C_n}.(\frac{p}{q})^n.f(0)$$

$$Since,\ f(x)\ is\ a\ \ p.m.f.\ so\ \sum_{x=0}^{n} f(x)=1.$$

$$or,\ f(0)+f(1)+f(2)+f(3)+....+f(n)=1$$

$$or,\ f(0)[1+n_{C_1}.\frac{p}{q}+n_{C_2}.\left(\frac{p}{q}\right)^2+n_{C_3}.\left(\frac{p}{q}\right)^3$$

$$+.....+n_{C_n}.\left(\frac{p}{q}\right)^n]=1$$

$$or,\ f(0)[1+\frac{p}{q}]^n=1$$

$$or,\ f(0) = q^n$$

$$So,\ f(1) = n_{C_1} p q^{n-1}, \quad f(2) = n_{C_2} p^2 q^{n-2},$$

$$\ldots\ldots\ldots\ldots f(n) = n_{C_n} p^n$$

$$\therefore\ f(x) = n_{C_x} p^x q^{n-x}, \quad x = 0,1,2,\ldots\ldots\ldots,n$$

Problem:

$$Mean\ and\ s\tan dard\ deviation\ of\ a\ binomial\ distribution\ are\ respectively\ 4\ and\ \sqrt{\frac{8}{3}}.\ Find\ the\ values\ of\ n\ and\ p.$$

Solution:

$$Since,\ X \sim Bin(n,p),\ so\ np = 4\ and\ npq = \frac{8}{3}.$$

$$\therefore q = {}^{8}\!/_{12} = {}^{2}\!/_{3},\ p = {}^{1}\!/_{3}\ and\ n = 12.$$

Mr. Pradip Panda, Statistics Department,
Serampore College

Problem:

$$If\ \ a\ random\ variable\ X\ \ follows\ binomial\ distri-$$

$$bution\ with\ mean\ \frac{5}{3}\ and\ P(X=1)=P(X=2),$$

$$find\ P(X\ is\ at\ most\ 1)\ and\ P(X\ is at\ least\ 1).$$

Solution: $If\ \ X \sim Bin(n,p)\ then\ mean\ np=5/3.$

$$P(X=1)=P(X=2)\ or,\ n.p.q^{n-1}=\frac{n(n-1)}{2}.p^2.q^{n-2}$$

$$or,\ 2.q=(n-1).p\quad or,\ 2.(1-p)=np-p$$

$$or,\ p = 2 - \frac{5}{3} = \frac{1}{3} \qquad \therefore q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$np = \frac{5}{3} \Rightarrow n = 5.$$

$$P(X \le 1) = P(X = 0) + P(X = 1)$$

$$= (\frac{2}{3})^5 + 5.\frac{1}{3}.(\frac{2}{3})^4 = 0.46$$

$$P(X \ge 1) = 1 - P(X = 0) = 1 - (\frac{2}{3})^5 = 0.87$$

Problem: $If\ \ X\ \ is\ a\ symmetric\ binomial$ $\text{var}\,iable\ with\ n = 36,\ calculate$ $E[X(X-1)].$

Solution: $Since,\ X\ \ is\ a\ symmetric\ binomial$ $\text{var}\,iable\ so\ \ p = 1/2\ and\ \ n = 36.$

$$E[X(X-1)] = E(X^2) - E(X).$$

$$= V(X) + E^2(X) - E(X).$$

$$= 36.\frac{1}{2}.\frac{1}{2} + 36^2.(\frac{1}{2})^2 - 36.\frac{1}{2}. \quad = 315$$

Problem:

$$For\ what\ value\ of\ \ p\ \text{var}\,iance\ will\ be\ \max imum,$$
$$if\ \ X \sim Bin(n, p)?$$

Solution:

$$Since,\ X \sim Bin(n, p)\ so\ V(X) = np(1-p)$$

$$\frac{dV(X)}{dp} = 0 \ \Rightarrow n - 2np = 0 \ \Rightarrow p = \frac{1}{2}$$

$$\left.\frac{dV(X)}{dp}\right|_{p=\frac{1}{2}} = -2n < 0.\ \therefore \text{var}\,iance\ is\ \max\ at\ \ p = \frac{1}{2}.$$

Problem: $If\ \ X \sim Bin(n, p)\ then\ \ prove\ that$

$$P(X\ \ is\ even) = \frac{1}{2}\left[1 + (q-p)^n\right],\ where\ \ p + q = 1.$$

Solution: $P(X\ \ is\ even)$

$$= P(X = 0) + P(X = 2) + P(X = 4) + \ldots\ldots.$$

$$= n_{C_0}.p^0.q^n + n_{C_2}.p^2.q^{n-2} + n_{C_4}.p^4.q^{n-4} + \ldots\ldots\ldots$$

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

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

Problem: $If\ \ X \sim Bin(n,p)\ then\ \ find\ \ the$

$$value\ of\ \ Cov(\frac{X}{n},\frac{n-X}{n}).$$

Solution:

$$Cov(\frac{X}{n},\frac{n-X}{n}) = -\frac{1}{n^2}.V(X).$$

$$= -\frac{n.p.q}{n^2} = -\frac{p.q}{n}$$

Problem:

Two persons toss a fair coin n times each. Show that the probability of their scoring the same number of heads is $2n_{C_n}.2^{-2n}$.

Solution:

Let X *be the number of heads for the first person and* Y *for the second person.*

$$P(X=Y)=P(X=0,Y=0)+P(X=1,Y=1)+\ldots$$

$$\ldots.+P(X=n,Y=n).$$

Mr. Pradip Panda, Statistics Department,
Serampore College

$$=[n_{C_0}.p^0.q^n]^2+[n_{C_1}.p^1.q^{n-1}]^2+......+[n_{C_n}p^n.q^0]^2$$

$$=[n_{C_0}]^2.[\frac{1}{2}]^{2.n}+[n_{C_1}]^2.[\frac{1}{2}]^{2.n}+......+[n_{C_n}]^2.[\frac{1}{2}]^{2.n}$$

$$=2n_{C_n}.2^{-2n}$$

SOME REAL LIFE SITUATIONS OF BINOMIAL DISTRIBUTION

- Failure and success in examination.
- Literate and illiterate people in a community.
- Arrival and non arrival of ships in the ports.
- Infection and non infection by diseases.
- Percentage of boys and girls in a given population.

Learn more……………