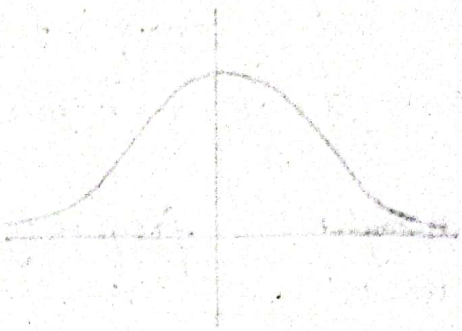
i) Multinomial Distribution
ii) Extreme value distribution
iii) Generalized Power Series Distribution
iv) Prior and posterior distribution
v) Compound Mixed type distribution
vi) Non-parametric test
⇓
No specific distribution
vii) Testing of Hypothesis



Random Variable:

The variable which takes different values generated by random experiment is known as random variable.

Random variables are denoted by X, Y, and Z.

Variables
- Quantitative
  - Discrete — Numbers — Count
  - Continuous — Interval — Measure
- Qualitative — Categorical

For example: if we ~~toss a coin~~ toss a coin three times, then total number of possible cases

$= 8$

no. of totals (events)

$= 2^3 = 8$

Sample space

	HH	HT	TH	TT
H	HHH	HHT	HTH	HTT
T	THH	THT	TTH	TTT

X = number of heads

X	$P(X)$
3	1/8
2	3/8
1	3/8
0	1/8
	$\Sigma P(X) = 1$

where, X is a random variable.

Weight	frequency	Probability
45-55	5	5/35
55-65		
65-75		
75-85		
85+		
	35	$\Sigma = 1$

Continuous

Mathematical Expectation (Average value)

Let, X be a discrete random variable which takes the values $x_1, x_2, \ldots, x_n$ with respective probabilities $P(x_1)$, $P(x_2), \ldots, P(x_n)$ then the mathematical expectation of random variable X is given by

$$E(X) = x_1 P(x_1) + x_2 P(x_2) + \ldots + x_n P(x_n)$$

$$= \Sigma x P(x)$$

Probability mass function. ; $1 \le x \le n$

If X is continuous random variable, then

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx \quad ; \quad -\infty < x < \infty$$

$f(x)$ = Probability density function

Suppose a dice is rolled once, then total number of possible cases = 6

X = values on the face of dice

X	$P(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
	$\Sigma P(X) = 1$

$$E(X) = 1 \cdot 1/6 + \cdots + 6 \cdot 1/6$$
$$= 1/6\,[1 + 2 + \cdots + 6]$$
$$= \frac{\Sigma X}{n}$$
$$= \bar{X}$$

Variance in terms of mathematical expectation

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

where, $E(X) = \Sigma X P(X)$

$E(X^2) = \Sigma X^2 P(X)$

$$\frac{\Sigma (x - \bar{x})^2}{n}$$

= Average of square of deviations

If X is continuous,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx$$

Second order moment about origin.

$$\frac{\Sigma x^2}{n} - \bar{x}^2$$
$$= \overline{x^2} - \bar{x}^2$$
$$= E(X^2) - [E(X)]^2$$
$$= E(X^2) - [E(X)]^2$$

x	$\bar{x}$	$y = (x - \bar{x})^2$
⋮	⋮	⋮

$$\bar{Y} = \Sigma y / n = \frac{\Sigma (x - \bar{x})^2}{n}$$

$V(X)$

$$E(Y) = E[X - \bar{X}]^2$$
$$= E[X - E(X)]^2$$

Covariance in terms of mathematical expectation

$$Cov(X, Y) = E\left[\{X - E(X)\}\{Y - E(Y)\}\right]$$

$$= E(XY) - E(X)E(Y)$$

$$r = \frac{Cov(X, Y)}{sd(X) * sd(Y)}$$

$$-1 \le r \le +1$$

$$Cov(X, Y) = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{n}$$

$$= \frac{\Sigma Z}{n}$$

$$= \bar{Z}$$

$$E(Z) = E\left[(X - \bar{X})(Y - \bar{Y})\right]$$

$$= E\left[X - E(X)\right]\left[Y - E(Y)\right]$$

$$= \frac{\Sigma XY}{n} - \bar{X}\bar{Y}$$

Moments

The average of various powers of deviations taken from chosen value is known as moments.

Mean | Any arbitrary value

eg: $\frac{\Sigma (x - a)^r}{n}$ ← Raw moment

If the deviations are taken from mean, then it is called central moments and if from any arbitrary value (assumed mean) then it is known as raw moments.

July 11'24

Central moments:

The rth order central moment is denoted by μ_r and is given by

$$\mu_r = \frac{1}{n} \Sigma (x - \bar{x})^r$$

First four central moments

$$\mu_1 = \frac{1}{n} \Sigma (x - \bar{x})^1 = 0$$

$$\mu_2 = \frac{1}{n} \Sigma (x - \bar{x})^2 = \text{Variance}$$

$$\mu_3 = \frac{1}{n} \Sigma (x - \bar{x})^3$$

$$\mu_4 = \frac{1}{n} \Sigma (x - \bar{x})^4$$

CamScanner

In terms of expectation,

$$\mu_r = E[X - E(X)]^r$$

Raw moments

The rth order raw moment about any arbitrary value 'a' is denoted by μ_r' and is given by:

$$\mu_r' = \frac{1}{n} \Sigma (x-a)^r$$

The rth raw moment about origin

$$\mu_r' = \frac{1}{n} \Sigma x^r \quad [\because a = 0]$$

In terms of expectation,

$$\mu_r' = E[X-a]^r$$

$$\mu_r' = E(X)^r \quad (\text{if } a = 0)$$

NOTE: $\mu_2 = \mu_2' - (\mu_1')^2$

Moment Generating Function

The moment generating function of random variable X is given by

$$M_X(t) = E[e^{tX}]$$

$$= \sum_x e^{tx} P(x) \quad \text{(For discrete)}$$

Pmf

$$= \int_x e^{tx} f(x)\, dx \quad \text{(For continuous)}$$

Pdf

NOTE: $M_X(t) = E\left[1 + \frac{tX}{1!} + \frac{t^2X^2}{2!} + \cdots + \frac{t^rX^r}{r!} + \cdots\right]$

$$= \left[1 + \frac{t}{1!} E(X) + \frac{t^2}{2!} E(X^2) + \cdots + \frac{t^r}{r!} E(X^r) + \cdots\right]$$

Expansion of e^x:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$E(aX) = a E(X)$

$\mu_1' = E(X) = \text{mean}$

$$\Rightarrow M_X(t) = \left[1 + \frac{t}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots + \frac{t^r}{r!}\mu_r' + \dots\right]$$

$\Rightarrow \mu_r'$ = coefficient of $\frac{t^r}{r!}$ in the expansion of $M_X(t)$.

$$\left.\frac{d M_X(t)}{dt}\right|_{t=0} = \left. 0 + \mu_1' + \frac{2t}{2!}\mu_2' + \dots + \frac{r t^{r-1}}{r!}\mu_r' + \dots \right|_{t=0}$$

$$= \mu_1'$$

$$= E(X)$$

Further,

$$\mu_r' = \left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0}$$

$$\therefore \mu_1' = \left.\frac{d M_X(t)}{dt}\right|_{t=0} = E(X)$$

$$\therefore \mu_2' = \left.\frac{d^2 M_X(t)}{dt^2}\right|_{t=0} = \underline{E(X^2)}$$

NOTE:

$$\boxed{\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'}$$

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

$$E[X^2] = E[X-0]^2$$

$$\mu_r = E[X-\bar{X}]^r$$

$$\mu_r' = E[X-a]^r$$

Put $a = 0$

r^{th} order raw moment about origin,

$$\mu_r' = E[X^r]$$

$$\mu_1' = E[X] = \text{mean}$$

$$\mu_2' = E[X^2] =$$

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

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

CamScanner

Characteristics Function

The characteristic function of random variable X is given by

$$\phi_x(t) = E\left[e^{itX}\right]$$

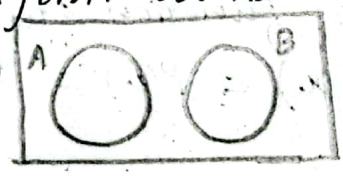
$$= \sum_x e^{itx} P(x) \quad [\text{ for discrete }]$$

$$= \int_x e^{itx} f(x)\, dx \quad [\text{ for continuous }]$$

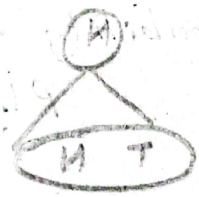
Generalized Power Series Distribution

i) Binomial Distribution
ii) Poisson distribution
iii) Negative binomial distribution

* Independent events : Occurence of first one does not influence second event.
* Mutually exclusive events : Disjoint events



A or A'



$n = 10$ → fixed (no. of trials)

* Events are independent
* Events are mutually exclusive
* Probability of success (p) is constant

$X =$ values

i) Binomial Distribution

Under the following conditions, the binomial distribution is applied:

a. Number of trials 'n' is fixed.
b. Events are statistically independent.
c. Events are mutually exclusive i.e. only two possible outcomes like success or failure, head or tail, either male or female etc.

CamScanner

d. The probability of success (p) is constant in each trial.

e. $p + q = 1$

Coin toss

(H) Success or Failure (T)

$1/2$ $1/2$

Product $\longrightarrow$ Defective or non-defective

0.01 0.99

n, p are parameters of the distribution.

<u>Definition</u>

Let, X be a discrete random variable which follows binomial distribution with parameters 'n' and 'p', then the probability mass function of the distribution is given by

$$P(X = x) = P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

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

$p + q = 1$

It gives the probability of getting 'x' successes out of 'n' independent trials

X = no. of success

p = probability of success in each trial

$q = 1 - p$ = probability of failure in each trial

<u>Example:</u>

Coin tossed 5 times, probability of getting 2 heads is: $P(X = 2) = {}^{5}C_{2}\,(0.5)^{2}\,(0.5)^{3}$

$n = 5$

$p = 1/2$, $q = 1/2$

CamScanner

Generalized Power Series Distribution

A discrete random variable X is said to follow a general power series distribution (GPSD) if its probability mass function is given by

$$P(X=x) = P(x) = \begin{cases} \dfrac{a_x \theta^x}{\cancel{f(\theta)(1+\theta)^n}} & ; x = 0, 1, 2, \ldots \\ & a_x > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

Power series

where, $\cancel{f(\theta)} = \sum_{x \in S^*} a_x \theta^x = f(\theta)$; $\theta \geq 0$

so that $f(\theta)$ is positive, finite and differentiable and 'S' is non-empty countable subset of non-negative integers.

Some special cases of GPSD

i) Binomial distribution:

In GPSD, take $\theta = \dfrac{p}{1-p}$ and $f(\theta) = (1+\theta)^n$ and $S = \{0, 1, 2, \ldots, n\}$

The probability mass function of GPSD is $P(X=x) = P(x) = \dfrac{a_x \theta^x}{f(\theta)}$

$$\boxed{(a+b)^n = \sum_{x=0}^{n} {}^nC_x b^x a^{n-x}}$$

where, $f(\theta) = \sum_{x \in S} a_x \theta^x \quad \ldots (i)$

Now, $f(\theta) = (1+\theta)^n = \sum_{x=0}^{n} {}^nC_x \theta^x \, 1^{n-x}$

$$= \sum_{x=0}^{n} {}^nC_x \theta^x \quad \ldots (ii)$$

From (i) and (ii), $a_x = {}^nC_x$

$$\therefore P(X=x) = \frac{{}^nC_x \left(\frac{p}{1-p}\right)^x}{\left(1 + (p/1-p)\right)^n} = \frac{{}^nC_x \, p^x (1-p)^{n-x}}{(1-p+p)^n}$$

* Sample space

CamScanner

$$= {}^nC_x\, p^x q^{n-x} \quad ; \quad x = 0, 1, 2, \ldots, n$$

This is the probability mass function of the binomial distribution.

July 13, 2024

$$(a+b)^n = \sum_{x=0}^{n} {}^nC_x\, b^x a^{n-x}$$

* Mean and Variance of binomial distribution

If $X \sim B(n,p)$ [i.e. X follows binomial distribution with parameters 'n' and 'p'], then

$$P(X = x) = {}^nC_x\, p^x q^{n-x} \quad ; \quad x = 0, 1, 2, \ldots, n$$

$$p + q = 1$$

<u>Mean</u>:

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

$$= \sum_{x=0}^{n} x\, {}^nC_x\, p^x q^{n-x}$$

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

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

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

$n! = n \times (n-1)!$

$${}^nC_r = \frac{n!}{(n-r)!\,r!}$$

$$= np \sum_{x=1}^{n} {}^{n-1}C_{x-1}\, p^{x-1} q^{(n-1)-(x-1)}$$

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

$$= np\,(q+p)^{n-1} \quad (\because p + q = 1)$$

$$\boxed{\therefore E(X) = np}$$

CamScanner

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

$$\Rightarrow Var(X) = E(X^2) - (np)^2 \quad \text{———— (i)}$$

Now, $E(X^2) = \sum_{x=0}^{n} x^2 \, P(X=x)$

$$= \sum_{x=0}^{n} x^2 \, {}^nC_x \, p^x q^{n-x}$$

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

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

$$\underbrace{\sum_{x=0}^{n} x \cdot \frac{n!}{(n-x)!\,x!} p^x q^{n-x}}_{np} \longleftarrow E(X) = np$$

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

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

$$= n(n-1)p^2 \sum_{x=2}^{n} {}^{n-2}C_{x-2} \, p^{x-2} q^{(n-2)-(x-2)} + np$$

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

$$\therefore E(X^2) = n(n-1)p^2 + np \qquad [\because p+q=1]$$

From (i), $Var(X) = n(n-1)p^2 + np - n^2p^2$

$$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2} = np(1-p)$$

$$\boxed{\therefore Var(X) = npq}$$

Hence, mean and variance of binomial distribution with ~~are~~ parameters n and p are respectively np and npq.

* Moment Generating function of binomial distribution

If $X \sim B(n,p)$ then $P(X=x) = {}^nC_x p^x q^{n-x}$; $x = 0, 1, 2, \dots, n$
$p+q=1$.

Now, moment generating function of X,

$$M_X(t) = E\left[e^{tX}\right]$$

$$= \sum_{x=0}^{n} e^{tx} P(x)$$

$$= \sum_{x=0}^{n} e^{tx} \, {}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nC_x (pe^t)^x q^{n-x}$$

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

Mean and Variance using MGF (mgf)

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'$$

$E[X-0]$

rth order raw moment about origin

We know that, $\mu_r' = \left.\dfrac{d^r M_X(t)}{dt^r}\right|_{t=0}$

If $X \sim B(n,p)$ then $M_X(t) = (q + pe^t)^n$

Now,

$$\mu_1' = E(X) = \left.\frac{d M_X(t)}{dt}\right|_{t=0} = n(q+pe^t)^{n-1} \cdot pe^t$$

Putting $t=0$,

$$= n(q+pe^0)^{n-1} \cdot pe^0$$

CamScanner

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

$$\boxed{\therefore E(X) = np} \qquad (\because p+q=1)$$

Further, $\mu_2' = E(X^2)$ = Second order raw moment about origin.

$$= \left.\frac{d^2 M_x(t)}{dt^2}\right|_{t=0}$$

$$= \left.\frac{d}{dt}\left[n(q+pe^t)^{n-1} pe^t\right]\right|_{t=0}$$

$$= np\left[e^t(n-1)(q+pe^t)^{n-2} pe^t + (q+pe^t)^{n-1} \cdot e^t\right]_{t=0}$$

$$= np\left[e^0(n-1)(q+pe^0)^{n-2} pe^0 + (q+pe^0)^{n-1} e^0\right]$$

$$= np\left[(n-1)p + 1\right]$$

$$= np(np - p + 1)$$

$$\therefore E(X^2) = (n^2p^2 - np^2 + np)$$

Now, $V(X) = E[X^2] - [E(X)]^2$

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

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(1-p)$$

$$\boxed{\therefore \mu_2 = npq}$$

CamScanner

Q2. *PROBLEMS*

Given,

$n = 7$ and $p = 0.2$, $q = 0.8$

(a) $P(X = 5)$

If $X \sim B(n, p)$ then $P(X = x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$; $x = 0, 1, \ldots, n$

$p + q = 1$

(a) $P(X = 5) = {}^{7}C_{5}\, 0.2^{5}\, 0.8^{2} = \underline{0.0043}$

(b) $P(X > 2) = 1 - P(X \leq 2)$

$\sum P(x) = 1$

$= 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - \left({}^{7}C_{0}\, 0.2^{0}\, 0.8^{7} + {}^{7}C_{1}\, 0.2^{1}\, 0.8^{6} + {}^{7}C_{2}\, 0.2^{2}\, 0.8^{5}\right)$

$= 1 - (0.2097 + 0.3670 + 0.2752)$

$= \underline{0.148}$

(c) $P(X \geq 3)$

(d) $P(X \geq 4)$

Q4.

$n = 6$

$P(X=3) = 0.2457$

$P(X=4) = 0.089$

$p = ?$

If $X \sim B(n, p)$ then, $P(X = x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$; $x = 0, 1, 2, \ldots, n$

$p + q = 1$

Then,

$$\frac{P(X=3)}{P(X=4)} = \frac{{}^{6}C_{3}\, p^{3}\, q^{6-3}}{{}^{6}C_{4}\, p^{4}\, q^{6-4}}$$

$$\Rightarrow \frac{0.2457}{0.089} = \frac{{}^6C_3\, p^3 q^3}{{}^6C_4\, p^4 q^2}$$

$$\Rightarrow 2.76 = \frac{20q}{15p} \quad \Rightarrow \frac{1-p}{p} = \frac{2.76 \times 15}{20}$$

$$\Rightarrow \frac{1-p}{p} = 2.07$$

$$\Rightarrow 1/p = 3.07$$

$$\therefore p = \underline{0.325}$$

July 16, 2024

Q5.

$n \rightarrow$ fixed

$p \rightarrow$ constant

Events are independent and mutually exclusive

$p+q=1$

p = probability of withdraw = 0.20

$q = 1-p = 1-0.20 = 0.80$

X = number of withdraw

$n = 18$

Here, $X \sim B(n, p)$, then $P(X=x) = {}^nC_x\, p^x q^{n-x}$; $x = 0, 1, \ldots, n$

$p+q=1$

(a) Probability that none will withdraw,

$$P(X=0) = {}^{18}C_0 (0.20)^0 (0.80)^{18-0}$$

$$= \underline{0.018}$$

(b) $P(X \geq 1) = 1 - P(X<1) = 1 - P(X=0)$

$$= 1 - 0.018$$

$$= \underline{0.982}$$

CamScanner

(c) $P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$= {}^{18}C_0 (0.2)^0 (0.8)^{18-0} + {}^{18}C_1 (0.2)^1 (0.8)^{18-1} + {}^{18}C_2 (0.2)^2 (0.8)^{18-2}$

$= 0.018 + 0.0810 + 0.1722$

$= \underline{0.2712}$

Assignment
Q6, Q7, Q13, Q16,
Q20, Q24

Q N10

$p = 0.05$

$q = 1 - 0.05 = 0.95$

$n = 3$ car that needs

X : No. of warranty repair in first 90 days.

Here, $X \sim B(n, p)$, so $P(X=x) = {}^nC_x \, p^x q^x$; $x = 0, 1, 2, \ldots, n$

$p + q = 1$

(a) $P(X=0) = {}^3C_0 (0.05)^0 (0.95)^{3-0}$

$= \underline{0.8573}$

(b) $P(X \ge 1) = 1 - P(X<1)$

$= 1 - P(X=0)$

$= 1 - 0.8573$

$= \underline{0.1427}$

(c) $P(X>1) = 1 - [P(X=0) + P(X=1)]$

$= 1 - (0.8573 + {}^3C_1 (0.05)^1 (0.95)^{3-1})$

$= 1 - (0.8573 + 0.1353)$

$= \underline{0.0074}$

CamScanner

QN14

(a) $n = 10$

$p = 0.3$

$q = 1 - p = 1 - 0.3 = 0.7$

X = number of liberals

Here, $X \sim B(n, p)$ then, $P(X = x) = {}^nC_x \, p^x q^{n-x}$; $x = 0, 1, 2, \ldots n$
$p + q = 1$

$$P(X = 4) = {}^{10}C_4 (0.3)^4 (0.7)^{10-4}$$
$$= 210 \times 0.0081 \times 0.1176$$
$$= \underline{0.200}$$

(b) X = number of conservative

$p = 0.55$, $q = 1 - p = 1 - 0.55 = 0.45$
$n = 10$

Here, $X \sim B(n, p)$, then $P(X = x) = {}^nC_x \, p^x q^{n-x}$; $x = 0, 1, 2, \ldots, n$
$p + q = 1$

$$\therefore P(X = 0) = {}^{10}C_0 (0.55)^0 (0.45)^{10-0}$$
$$= 1 \times 1 \times 0.00034$$
$$= \underline{0.00034}$$

(c) X = number of middle of the road

$p = 0.15$

$q = 1 - p = 1 - 0.15 = 0.85$
$n = 10$

Here, $X \sim B(n, p)$, then $P(X = x) = {}^nC_x \, p^x q^{n-x}$; $x = 0, 1, 2, \ldots n$,

CamScanner

$\therefore P(X=2) = {}^{10}C_2 (0.15)^2 (0.85)^{10-2}$

$= \underline{0.2758}$

(d) X = number of liberals

$P(X \geq 8) = ?$

$P(X \geq 8) = P(X=8) + P(X=9) + P(X=10)$

$= {}^{10}C_8 (0.3)^8 (0.7)^{10-8} + {}^{10}C_9 (0.3)^9 (0.7)^{10-9}$

$+ {}^{10}C_{10} (0.3)^{10} (0.7)^{10-10}$

$=$

QN 20

Here, $np = 0.4$ —— (i)

$\sqrt{npq} = 0.6$

$\Rightarrow npq = 0.36$ —— (ii)

Dividing (i) by (ii)

$\frac{1}{q} = \frac{0.4}{0.36}$

$\therefore q = 0.9$

$\therefore p = 0.1$ $\therefore n = 0.4/0.1 = 4$

CamScanner

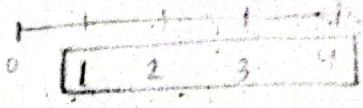
If $X \sim B(n, p)$, then $P(X = x) = {}^nC_x \, p^x q^{n-x}$; $x = 0, 1, 2, \ldots, n$
$p + q = 1$

$\therefore P(X \geq 1) = 1 - P(X = 0)$

$= 1 - {}^4C_0 (0.1)^0 (0.9)^{4-0}$

$= 1 - 1 \times 1 \times 0.6561$

$= \underline{0.3439}$



Q 23

$p = 1/4 = 0.25$

$q = 1 - 0.25 = 0.75$

$n = 5$

X = number of correct questions

Here, $X \sim B(n, p)$, then $P(X = x) = {}^nC_x \, p^x q^{n-x}$; $x = 0, 1, \ldots, n$
$p + q = 1$

(a) $P(X = 5) = {}^5C_5 (0.25)^5 (0.75)^{5-5}$

$= 1 \times 0.00097 \times 1$

$= \cancel{0.00485}$ $\underline{0.00097}$

(b) $P(X \geq 4) = P(X = 4) + P(X = 5)$

$= {}^5C_4 (0.25)^4 (0.75)^{5-4} + {}^5C_5 (0.25)^5 (0.75)^{5-5}$

$= 0.01464 + \cancel{0.00485}$ 0.00097

$= \underline{0.01561}$

CamScanner

POISSON'S DISTRIBUTION (Use for rare events)

λ = Average number of calls per hour $\qquad X = 0, 1, 2, \ldots$

$= 5$ (say)

X = Number of calls per hour

$P(X \geq 1) = ?$

$= 1 - P(X < 1)$

$= 1 - P(X = 0)$

$$\frac{e^{-\lambda} \lambda^{x}}{x!}$$

$n \longrightarrow \infty$

$p \longrightarrow 0$

$\lambda = np$

$X = 0, 1, \ldots$

Let X be a discrete random variable which follows Poisson's distribution with parameter 'λ' then, the probability mass function (pmf) of the distribution is given by

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \; ; \quad x = 0, 1, 2, \ldots \quad \lambda > 0$$

where, λ = average value

Poisson distribution using GPSD

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

where, $f(\theta) = \sum_{x \in S} a_x \theta^x$

$\theta \geq 0$

$x \in S$

$\downarrow \; 0, 1, 2, \ldots, \infty$

$x = ?$

$f(\theta) = ?$

$S = ?$

In GPSD, put $\theta = \lambda$, $f(\theta) = e^{\theta}$
and $S = \{0, 1, 2, \ldots, \infty\}$

Now, $f(\theta) = e^{\theta} = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \cdots$

$$f(\theta) = \sum_{x=0}^{\infty} \frac{\theta^x}{x!} \quad \text{——— (i)}$$

Since, $f(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$ ——— (ii)

From (i) and (ii), $a_x = \frac{1}{x!}$

The probability mass function of Generalized Power Series Distribution is

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

$$= \frac{\frac{1}{x!} \lambda^x}{e^{\lambda}}$$

$$\boxed{\therefore P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}} \; ; \quad x = 0, 1, 2, \ldots \quad \lambda > 0.$$

This is the probability mass function of Poisson's distribution with parameter λ.

Mean and Variance of the distribution

If $X \sim P(\lambda)$, then $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$; $x = 0, 1, \ldots$ $\lambda > 0$

<u>Mean:</u>

$$E(X) = \sum_{x=0}^{\infty} x P(x) = \sum_{x=0}^{\infty} x \left\{ \frac{e^{-\lambda} \lambda^x}{x!} \right\} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{x \cdot \lambda^x}{x(x-1)!}$$

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

CS CamScanner

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

$$= \lambda e^{-\lambda} e^{\lambda}$$

$$\boxed{\therefore E(X) = \lambda}$$

July 18, '24

Q24) No. of trials, $(n) = 5$

[7, 11 or 12] (If a pair of dice is rolled)

p = probability that a player gets audited

No. of possible cases $= 6^2$

$= 36$

$$\therefore p = \frac{9}{36} = \frac{1}{4} = 0.25$$

$q = 1 - p = 1 - 0.25 = 0.75$

X = no. of times she gets audited

Here, $X \sim B(n, p)$, then $P(X = x) = {}^nC_x \, p^x \, q^{n-x}$; $x = 0, 1, 2, \ldots$

$p + q = 1$

$$P(X \geq 1) = 1 - P(X < 1)$$

$$= 1 - P(X = 0)$$

$$= 1 - {}^5C_0 (0.25)^0 (0.75)^{5-0}$$

$$= \underline{0.7627}$$

Variance of Poisson's distribution

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

$$= E(X^2) - \lambda^2 \quad \text{———} \quad (i)$$

Now,

$$E(X^2) = \sum_{x=0}^{\infty} x^2 P(x)$$

$$= \sum_{x=0}^{\infty} [x(x-1) + x] P(x)$$

$$= \sum_{x=0}^{\infty} x(x-1) P(x) + \sum_{x=0}^{\infty} x P(x)$$

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

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

$$= e^{-\lambda} \cdot \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$$

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

$$= \lambda^2 e^{-\lambda} \cdot e^{\lambda} + \lambda$$

$$= \lambda^2 + \lambda$$

From (i),

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

$$= \lambda^2 + \lambda - \lambda^2$$

$$\boxed{\therefore V(X) = \lambda}$$

CamScanner

Moment Generating Function

If $X \sim P(\lambda)$, then $P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$; $x = 0, 1, 2, \ldots$
$\lambda > 0$

Now, mgf of X,

$$M_X(t) = E\left[e^{tX}\right]$$

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

$$= \sum_{x=0}^{\infty} e^{tx} \left\{ \frac{e^{-\lambda}\lambda^{x}}{x!} \right\}$$

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

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

$$= e^{-\lambda} \left(e^{\lambda e^{t}} \right)$$

$$= \cancel{e^{\lambda (e^{t}}}$$

$$\therefore M_X(t) = e^{\lambda(e^{t}-1)} \quad \text{———} \quad (i)$$

Mean and variance using mgf: $e^{\lambda(e^{t}-1)}$

If $X \sim P(\lambda)$, then $M_X(t) = e^{\lambda(e^{t}-1)}$

Since $\mu_r' = \left.\dfrac{d^{r} M_X(t)}{dt^{r}}\right|_{t=0}$

Mean: $\mu_1' = E(X) = \left.\dfrac{d\, M_X(t)}{dt}\right|_{t=0}$

$$= \left.\frac{d}{dt}\left(e^{\lambda(e^{t}-1)} \right)\right|_{t=0}$$

$$= \cancel{e^{\lambda \cdot e^{t-1}}} = e^{\lambda(e^{t}-1)} \cdot \lambda \cdot e^{t} \Big|_{t=0}$$

$$= e^{\lambda(e^{0}-1)} \cdot \lambda e^{0}$$

$$= e^{\lambda \times 0} \cdot \lambda e^{0}$$

$$= 1 \cdot \lambda \cdot 1$$

$$\boxed{\therefore \text{Mean } (\mu_1') = \lambda}$$

<u>Variance</u> : $V(X) = E(X^2) - [E(X)]^2$

$$\mu_2' = E(X^2) = \left.\frac{d^2 M_X(t)}{dt^2}\right|_{t=0}$$

$$\frac{d^2 M_X(t)}{dt^2} = \frac{d}{dt}\left(e^{\lambda(e^{t}-1)} \cdot \lambda e^{t}\right)$$

$$= \lambda \frac{d}{dt}\left(e^{\lambda(e^{t}-1)} \cdot e^{t}\right)$$

$$= \lambda \left[e^{\lambda(e^{t}-1)} \cdot \frac{d}{dt}(e^{t}) + e^{t} \cdot \frac{d}{dt}\left(e^{\lambda(e^{t}-1)}\right)\right]$$

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

$$\therefore E(X^2) = \left.\frac{d^2 M_X(t)}{dt^2}\right|_{t=0} = \lambda\left[e^{\lambda(e^{0}-1)} \cdot e^{0} + e^{0}\left\{e^{\lambda(e^{0}-1)} \cdot \lambda e^{0}\right\}\right]$$

$$= \lambda\left[e^{\lambda \times 0} \times 1 + 1 \times e^{\lambda \times 0} \cdot \lambda \times 1\right]$$

$$= \lambda(1+\lambda)$$

$$= \lambda^2 + \lambda$$

$$\therefore V(X) = E(X^2) - [E(X)]^2 = \mu_2' - (\mu_1')^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$\boxed{\therefore V(X) = \lambda}$$

CamScanner

PROBLEMS

Q1. $\lambda = 4.2$

ⓐ $P(X \geq 5)$

$= 1 - P(X < 5)$ $\left[\because P(X = x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} ; \quad x = 0, 1, 2, \ldots \quad \lambda > 0\right]$

$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)]$

$$= 1 - \left[\frac{e^{-4.2}\, 4.2^{0}}{0!} + \frac{e^{-4.2}\, 4.2^{1}}{1!} + \frac{e^{-4.2}\, 4.2^{2}}{2!} + \frac{e^{-4.2}\, 4.2^{3}}{3!} + \frac{e^{-4.2}\, 4.2^{4}}{4!}\right]$$

$= 1 - (e^{-4.2} / e^{4.2})$

$= 1 - 1/e$

$= 1 - e^{-4.2}[1 + 4.2 + 8.82 + 12.34 + 12.96]$

$= 1 - 0.589$

$= \underline{0.410}$

ⓑ $P(X < 2) = P(X=0) + P(X=1)$

$= e^{-4.2}[1 + 4.2]$

$= \underline{0.0779}$

CS CamScanner

Q N3) X = No. of price hikes every three years

λ = average number of price hikes every 3 years

= 4 (every three years)

Here, $X \sim P(\lambda)$ so, $P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$; $x = 0, 1, 2, \ldots$ $\lambda > 0$

X = no. of successes

p = prob. of success

(a) $P(X=0) = \dfrac{e^{-4}4^{0}}{0!} = e^{-4} = \underline{0.0183}$

(b) $P(X=2)$

(c) $P(X=4)$

(d) $P(X \geq 5) = 1 - P(X < 5)$

$=$

CS CamScanner

Q5) X = Number of claims per hour

λ = Average number of claims per hour

$= 3.1$

Here, $X \sim P(\lambda)$ so, $P(X = x) = \dfrac{e^{-\lambda} \lambda^{x}}{x!}$; $x = 0, 1, 2, \ldots$

$\lambda > 0$

(a) $P(X < 3) = P(X=0) + P(X=1) + P(X=2)$

$$= \frac{e^{-3.1}\, 3.1^{0}}{0!} + \frac{e^{-3.1}\, 3.1^{1}}{1!} + \frac{e^{-3.1}\, (3.1)^{2}}{2!}$$

$$= e^{-3.1}\left(1 + 3.1 + \cancel{4.80}\right)$$

$$= e^{-3.1} * \cancel{8.9}$$

$$= 0.045 * \cancel{8.9}\ 9.22$$

$$= \cancel{0.400}\ \underline{0.4149}$$

$$= e^{-3.1}\left(1 + 3.1 + 4.8\right)$$

$$= e^{-3.1} * 8.9$$

$$= \underline{0.400}$$

CamScanner

(b) $P(X=3) = \dfrac{e \quad 3.1'}{3!} = \underline{0.224}$

(c) $P(X \geq 3) = 1 - P(X < 3)$

$= 1 - 0.400$

$= \underline{0.600}$

(d) $P(X > 3) = 1 - P(X \leq 3)$

$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$

$= 1 - [P(X<3) + P(X=3)]$

$= 1 - [0.401 + 0.224]$

$= \underline{0.375}$

<u>Q7</u>

λ = Average number of arrivals in one minute

$= 6$

for (e), (f)

$\lambda = 12$

= Average number of arrivals in two minutes

X = Number of arrivals in one minute.

CS CamScanner

Poisson approximation to Binomial

The binomial distribution reduces to Poisson distribution under the following conditions:

(i) If no. of trials 'n' is very large i.e. $n \to \infty$ $(n > 20)$

(ii) if probability of success 'p' is very small i.e. $p \to 0$ $(p \leq 0.05)$

$\therefore \lambda = np$

$$\Rightarrow P(X = x) = \frac{e^{-np} (np)^x}{x!}$$

Poisson approximation to Binomial

The binomial distribution reduces to Poisson distribution under the following conditions:

(i) If no. of trials 'n' is very large i.e. $n \to \infty$ $(n > 20)$

(ii) if probability of success 'p' is very small i.e. $p \to 0$ $(p \leq 0.05)$

$$\therefore \lambda = np$$

$$\Rightarrow P(X = x) = \frac{e^{-np} (np)^x}{x!}$$

July 20, 2024

Q16) $n = 30, \; p = 0.04$

Using Poisson's approximation to Binomial, we get

$$\lambda = np$$
$$= 30 * 0.04$$
$$= \underline{1.2}$$

$$\therefore P(X = x) = \frac{e^{-np} (np)^x}{x!}$$

(a) $$P(X = 25) = \frac{e^{-1.2} (1.2)^{25}}{25!}$$

$$= 0$$

(b) $$P(X = 3) = \frac{e^{-1.2} (1.2)^3}{3!} = \underline{0.0867}$$

CamScanner

(c) $P(X=5) = \dfrac{e^{-1.2}(1.2)^5}{5!}$

$= \underline{0.00624}$

<u>Q17.</u> X = number of non-marital disputes

$p = 1 - 0.96$

$= 0.04$

$n = 80$

Here, $n > 20$ & $p \leq 0.05$,

Thus, using Poisson's approximation to binomial, we get:

$\lambda = np$

$= 80 \times 0.04 = 3.2$

$\therefore P(X=x) = \dfrac{e^{-np}(np)^x}{x!}$

(a) $P(X=7) = \dfrac{e^{-3.2} \times (3.2)^7}{7!}$

$= \dfrac{0.04076 * 3435.97383}{5040}$

$= \underline{0.02778}$

Q20.

$p = 1/400 = 0.0025$

$n = 100$

$X =$ no. of defective

Using Poisson's approximation to binomial,

$\lambda = np$

$= 0.0025 \times 100$

$= 0.25$

$\therefore P(X = x) = \dfrac{e^{-np} (np)^x}{x!}$

(a) $\therefore P(X = 0) = \dfrac{e^{-0.25} (0.25)^0}{0!}$

$= \underline{0.7788}$

(b) $P(X < 2) = P(X \leq 1) \quad P(X=0) + P(X=1)$

$= 0.7788 + \dfrac{e^{-0.25} (0.25)^1}{1!}$

$= 0.7788 + 0.1947$

$= 0.9735$

(c) $P(X \geq 1) = 1 - P(X = 0)$

$= 1 - 0.7788$

$= \underline{0.2212}$

(d) $P(X \geq 3) = 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - \left[0.9735 + \frac{e^{-0.25}(0.25)^2}{2!}\right]$

$= 1 - (0.9735 + 0.0243)$

$= \underline{0.0022}$

Assignment

#9 4, 7, 8, 13, 18, 19

Negative Binomial Distribution

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

Introduction: If a trial is repeated independently till fixed number of success occurs with probability 'p' which is constant for each trial, then such an experiment follows Negative Binomial Distribution (NBD).

$P(A \cap B)$ last trial

$= {}^{x+k-1}C_{k-1} \, p^{k-1} q^x \cdot p$

$= {}^{x+k-1}C_{k-1} \, p^k q^x$,

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

$k > 0$

${}^{x+k-1}C_x$

k = fixed number of successes

p = probability of success in each trial

${}^nC_x \, p^x q^{n-x}$

$P(A \cap B) = \underline{P(A) \cdot P(B)}$

Independent

$x + k$ = Total number of trials

$\downarrow$

no. of failures preceding 'k' success

$(x + k - 1)$ = trials

$(k - 1)$ = successes

Derivation of NBD

Consider a binomial experiment consists of 'n' trials, & the probability of success (p) is constant, for each trial. Let there be 'x' failures preceding the kth success in $x+k = n$ trials. Here $x+k$ trials are required to produce 'k' successes (which is fixed) and X is a random variable.

Now, we need to find the probability the kth success occurs in $x+k$ trials, which is the probability that exactly 'x' failures preceding the kth success in $x+k$ trials.

Here, the last trial must be a success with probability 'p' and the remaining $(x+k-1)$ trials, there are $(k-1)$ successes and whose probability is given by ${}^{x+k-1}C_{k-1}\, p^{k-1}\, q^{x}$.

Now, by multiplication theorem of probability, the probability of 'x' failures preceding the kth success in $(x+k)$ trials is given by

$$\left\{ {}^{x+k-1}C_{k-1}\, p^{k-1}\, q^{x} \right\} p$$

$$= {}^{x+k-1}C_{k-1}\, p^{k}\, q^{x}$$

This is the pmf of negative binomial distribution (NBD) with parameters 'k' and 'p'.

If $X \sim NB(k, p)$, then the probability mass function of the distribution is given by

$$P(X = x) = {}^{x+k-1}C_{k-1}\, p^{k}\, q^{x} \; ; \quad x = 0, 1, 2, \ldots$$

$$k > 0$$

$$p + q = 1$$

CamScanner

NBD as special case of GPSD

In GPSD, $\theta = \dfrac{q/p}{1+q/p}$, $f(\theta) = (1-\theta)^{-K}$

$S = 0, 1, 2, \ldots, \infty$

Now,

$$(1-\theta)^{-K} = \sum_{x=0}^{\infty} {}^{x+K-1}C_x \, \theta^x \quad \text{——— (i)}$$

$$f(\theta) = \sum_{x=0}^{\infty} a_x \theta^x \quad \text{——— (ii)}$$

From (i) and (ii),

$$a_x = {}^{x+K-1}C_x$$

Since, $P(X=x) = \dfrac{a_x \theta^x}{f(\theta)}$

$$\Rightarrow P(X=x) = \frac{{}^{x+K-1}C_x \left(\frac{q/p}{1+q/p}\right)^x}{\left[1 - \left(\frac{q/p}{1+q/p}\right)\right]^{-K}}$$

$$\Rightarrow P(X=x) = \frac{{}^{x+K-1}C_x \left(\frac{q/p}{1+q/p}\right)^x}{\left(\frac{1}{1+q/p}\right)^{-K}}$$

$$\Rightarrow P(X=x) = \frac{{}^{x+K-1}C_x \cdot \frac{q^x}{(p+q)^x}}{\left(\frac{p}{p+q}\right)^{-K}}$$

$$\Rightarrow P(X=x) = {}^{x+K-1}C_x \, q^x \cdot p^K$$

$$\boxed{\therefore P(X=x) = {}^{x+K-1}C_x \, p^K q^x}$$

This is the probability mass function of negative binomial distribution

$P(X=x) = \dfrac{a_x \theta^x}{f(\theta)}$

$f(\theta) = \sum_{x \in S} a_x \theta^x$; $\theta > 0$

$\theta = \dfrac{q/p}{1+q/p}$

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

$S = \{0, 1, 2, \ldots, \infty\}$

CS CamScanner

with parameters k and p. Hence, NBD is the special case of GPSD.

July 23, 2024

Mean and Variance of Negative Binomial Distribution.

If $X \sim NB(p, k)$ then, the pmf of the distribution,

$$P(X = x) = {}^{x+k-1}C_x \, p^k \cdot q^x \; ; \quad x = 0, 1, 2, \ldots$$
$$p + q = 1$$
$$k > 0$$

$$(1-a)^{-n} = \sum_{x=0}^{\infty} {}^{x+n-1}C_x \, a^x$$
$${}^{x+k-1}C_x = {}^{x+k-1}C_{x-1}$$

Mean, $E(X) = \sum_{x=0}^{\infty} x \, P(x)$

$$= \sum_{x=0}^{\infty} x \left\{ {}^{x+k-1}C_x \, p^k q^x \right\}$$

$$= p^k \sum_{x=0}^{\infty} x \frac{(x+k-1)!}{(\cancel{x}+k-1-\cancel{x})! \, x!} q^x$$

$$= p^k \sum_{x=0}^{\infty} \cancel{x} \frac{(x+k-1)!}{(\cancel{x}+k-1-\cancel{x})! \, \cancel{x}(x-1)!} q^x$$

$$= kp^k q \sum_{x=1}^{\infty} \frac{(x+k-1)!}{k! \, (x-1)!} q^{x-1}$$

$$= kp^k q \sum_{x=1}^{\infty} {}^{x+k-1}C_{x-1} \, q^{x-1}$$

$$= kp^k q \boxed{\sum_{(x-1)=0}^{\infty} {}^{(x-1)+(k+1)-1}C_{x-1} \, q^{x-1}}$$

$$= kp^k q \, (1-q)^{-(k+1)}$$

$$= kp^k q \, p^{-k-1}$$

$$= k \cancel{p^k} q \cdot \frac{1}{\cancel{p^k} \cdot p} = \frac{kq}{p}$$

$$\boxed{\therefore E(X) = \frac{kq}{p}}$$

CamScanner

Variance of NBD

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

$$= E(X^2) - \left(\frac{kq}{p}\right)^2 \quad \text{——— (i)}$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 P(x)$$

$$= \sum_{x=0}^{\infty} [x(x-1) + x] P(x)$$

$$= \sum_{x=0}^{\infty} x(x-1) P(x) + \sum_{x=0}^{\infty} x P(x)$$

$$= \sum_{x=0}^{\infty} x(x-1) \; {}^{x+k-1}C_x \, p^k q^x + \left(\frac{kq}{p}\right)$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{(x+k-1)!}{(\cancel{x}+k-1-\cancel{x})!\, x!} p^k q^x + \left(\frac{kq}{p}\right)$$

$$= p^k q^2 \sum_{x=0}^{\infty} \cancel{x(x-1)} \frac{(x+k-1)!}{(k-1)!\, \cancel{x(x-1)}(x-2)!} q^{x-2} + \left(\frac{kq}{p}\right)$$

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

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

$$= p^k q^2 k(k+1) \sum_{(x-2)=0}^{\infty} {}^{(x-2)+(k+2)-1}C_{x-2} \, q^{x-2} + \left(\frac{kq}{p}\right)$$

$$= p^k q^2 k(k+1) (1-q)^{-(k+2)} + (kq/p)$$

$$= p^k q^2 k(k+1) (p)^{-k+2} + kq/p = \frac{p^k q^2 k(k+1)}{p^k \cdot p^2} = \frac{q^2}{p^2} k(k+1) + kq/p$$

From ①,

$$Var(X) = \frac{q^2}{p^2} K(K+1) + \frac{kq}{p} - \frac{k^2 q^2}{p^2}$$

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

$$= kq/p \left(\frac{q}{p} + 1 \right)$$

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

$$= \frac{kq}{p} \times \frac{1}{p}$$

$$\boxed{\therefore Var(X) = \frac{kq}{p^2}}$$

Moment generating Function (Mgf)

If $X \sim NB(K, p)$, then $P(X = x) = {}^{x+k-1}C_x \, p^k q^x$; $x = 0, 1, 2, \ldots$, $p+q=1$, $K>0$

Now, mgf of X,

$$M_x(t) = E[e^{tX}]$$

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

$$= \sum_{x=0}^{\infty} e^{tx} \left\{ {}^{x+k-1}C_x \, p^k q^x \right\}$$

$$\Rightarrow M_x(t) = \left[\sum_{x=0}^{\infty} {}^{x+k-1}C_x \, (qe^t)^x\right] p^K$$

$$= p^K (1 - qe^t)^{-K}$$

$$\therefore M_x(t) = \frac{p^K}{(1-qe^t)^K}$$

This is the mgf of negative binomial distribution with parameters K and p.

$$\mu_1' = \text{Mean} = E(X) = \left.\frac{d\, M_x(t)}{dt}\right|_{t=0}$$

$$= \left.\frac{d}{dt}\left[p^K/(1-qe^t)^K\right]\right|_{t=0}$$

$$= \left[p^K \cdot (-K)(1-qe^t)^{-K-1} \cdot \frac{d}{dt}(1-qe^t)\right]_{t=0}$$

$$= -p^K K \cdot (1-qe^t)^{-K-1} \left\{-qe^t\right\}\Big|_{t=0}$$

$$= Kp^K (1-qe^t)^{-K-1} \, qe^t \Big|_{t=0}$$

$$= Kp^K (1-qe^0)^{-K-1} \cdot qe^0$$

$$= Kp^K (1-q)^{-K-1} \cdot q$$

$$= Kp^K q \, p^{-K-1} = \frac{K\cancel{p^K} q}{\cancel{p^K} \cdot p} = Kq/p$$

CamScanner

$$\mu_2' = \frac{d^2 M_x(t)}{dt^2} = kqp^k\left[e^t\left\{(-k-1)(1-qe^t)^{-k-2}(-qe^t)\right\} + (1-qe^t)^{-k-1}e^t\right]$$

$$\mu_2' = kqp^k\left[(-k-1)(1-q)^{-k-2}(-q) + p^{-k-1}\right]$$

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

$$= \frac{kq^2(k+1)}{p^2} + \frac{kq}{p}$$

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

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

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

$$= \frac{kq}{p}\left(q/p + 1\right)$$

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

$$\boxed{\therefore Var(X) = kq/p^2}$$

Q. The probability of hitting the target at any trial 0.2. If a shooter aims at a target, find the probability that fifth fire is second hit.

$p = 0.2$

$x + k = 5$

$k = 2$

$\Rightarrow x + 2 = 5$

$\therefore x = 3$

$P(X = 3)$

Given, p = probability of hitting a target in any trial $= 0.2$

x = number of mis hits

$k = 2$

$x + k = 5$

$\Rightarrow x = 3$

Here, $X \sim NBD(p, k)$ so $P(X = x) = {}^{x+k-1}C_x \, p^k q^x$; $x = 0, 1, 2, \ldots$

$p + q = 1$

$k > 0$

Now, probability that the fifth fire is second hit, $P(X = 3) = {}^{3+2-1}C_3 (0.2)^2 (0.8)^3$

$= {}^4C_3 (0.2)^2 (0.8)^3$

$= 0.08192$

Assignment

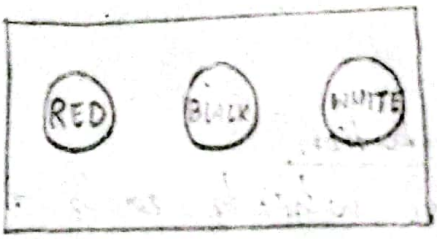
Q1. A boy is throwing stones at a target. If the probability of hitting the target at any trial is 0.5, what is the prob. that

10th throw is 5th hit?

Q2. The prob. that a striker will make a goal in a football mate is 0.3. What is the probability that he'll score fourth goal in 5th shot?

July 25, 2024

— MULTINOMIAL DISTRIBUTION —



At least 3 possible outcomes
→ with replacement

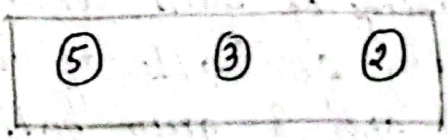
$$P(x_1, x_2, \dots, x_K) = \frac{n!}{x_1!\, x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K}$$

$$\sum_{i=1}^{K} x_i = n$$

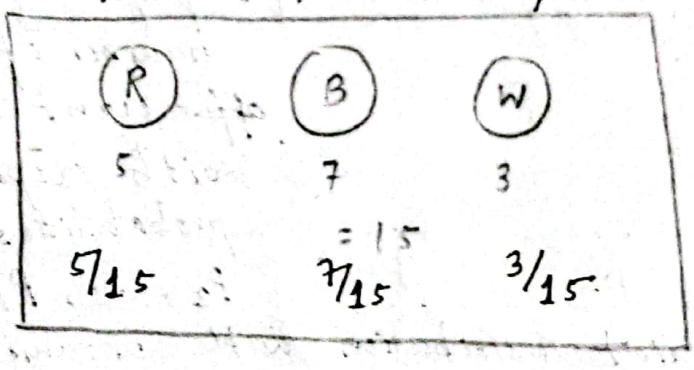
$$p_1 + p_2 + \dots + p_K = \sum_{i=1}^{K} p_i = 1$$

$n = 10$

$P(X_1 = 5, X_2 = 3, X_3 = 2)$



No. of trials = 5 using w/ replacement



$$P(X_1 = 2, X_2 = 2, X_3 = 1) = \frac{5!}{(2!)(2!)(1!)} \left(\frac{5}{15}\right)^2 \left(\frac{7}{15}\right)^2 \left(\frac{3}{15}\right)^1$$

CamScanner

Introduction

An experiment is said to be a multinomial ~~distribution~~ experiment if

i) the experiment consists of a fixed number of trials 'n'.

ii) ~~for~~ each trial there are $K(\geq 3)$ possible outcomes.

iii) Trials are statistically independent

iv) The probability of each outcome remains the same from trial to trial.

Examples of the multinomial distribution

(i) No. of throws of a fair dice in which each throw can result six different outcomes.

(ii) Number of selection or drawings of balls at random with replacement from a box containing 20 balls of which 2 are white, 4 are black, 6 Red and 8 green balls.

Definition: The random variable $X = (x_1, x_2, \ldots, x_K)$ denoting the outcome of 'n' trials where x_i = frequency or number of outcomes ~~of for~~ event E_i with respective probabilities (P_i) $P_1, P_2, \ldots, P_K$.

is said to have multinomial distribution with parameters $(n, P_1, P_2, \ldots, P_K)$, ~~then the~~ if its pmf is given by

$$P(X = x) = P(X = x_1, x_2, \ldots, x_K)$$

$$\boxed{= \frac{n!}{x_1!\, x_2! \ldots x_K!} P_1^{x_1} P_2^{x_2} \cdots P_K^{x_K}}$$

where, $n = x_1 + x_2 + \cdots + x_K = \sum_{i=1}^{K} x_i$

and $p = P_1 + P_2 + \cdots + P_K = \sum_{i=1}^{K} P_i$

CS CamScanner

$$\left[\sum_x P(x_1, x_2, \dots, x_K) = \sum_x \left[\frac{n!}{x_1!\, x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K}\right] = (p_1+p_2+\dots+p_K)^n\right.$$
$$= 1$$

Reduction of multinomial distribution when $k=2$

when $k=2$, then the pmf of multinomial distribution becomes

$$P(x_1, x_2) = \frac{n!}{x_1!\, x_2!} p_1^{x_1} p_2^{x_2}$$

where, $n = x_1 + x_2$ and $p_1 + p_2 = 1$

$\Rightarrow x_2 = (n - x_1)$ $\Rightarrow p_2 = (1 - p_1)$

$$P(x_1, n-x_1) = \frac{n!}{x_1!\,(n-x_1)!} p_1^{x_1} (1-p_1)^{(n-x_1)}.$$

Consider $x_1 = x$ and $p_1 = p$

$$P(x, n-x) = \frac{n!}{x!\,(n-x)!} p^x (1-p_1)^{n-x}$$

$$\Rightarrow \boxed{P(x) = {}^nC_x\, p^x (1-p)^{n-x}}$$

This is the pmf of binomial distribution with parameters 'n' and 'p'.

Hence, when $k=2$, the multinomial distribution reduces to binomial distribution with parameters 'n' and 'p'.

Moment generating function of Multinomial distribution

If $X \sim M_K(n, p_1, p_2, \dots, p_K)$ then pmf

$$P(x_1, x_2, \dots, x_K) = \frac{n!}{x_1!\, x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K}$$

Now, the moment generating function,

$$M_X(t) = M_{(X_1, X_2, \dots, X_K)}(t_1, t_2, \dots, t_K)$$

CamScanner

$$= E\left[e^{t_1 X_1} e^{t_2 X_2} \cdots e^{t_K X_K} \right]$$

$$= \sum_x \left(e^{t_1 x_1}\right) \cdots \left(e^{t_K x_K}\right) \frac{n!}{x_1!\, x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K}$$

$$= \sum_x \left(p_1 e^{t_1}\right)^{x_1} \cdots \left(p_K e^{t_K}\right)^{x_K} \frac{n!}{x_1!\, x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K}$$

$$\cancel{= p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_K e^{t_K}}$$

$$= \left(p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_K e^{t_K}\right)^n$$

$\because \sum_x \frac{n!}{x_1!\, x_2! \cdots x_K!} p_1^{x_1} \cdots p_K^{x_K} = (p_1 + p_2 + \cdots + p_K)^n$

If $t_1 \neq 0$ and $t_i = 0$ for $i = 2, 3, \ldots, k$, we get:

$$M_X(t) = M_{X_1, X_2, \ldots X_K}(t_1, 0, \ldots, 0)$$

$$= \left(p_1 e^{t_1} + \underbrace{p_2 + \cdots + p_K}\right)^n$$

$$= \left(p_1 e^{t_1} + (1 - p_1)\right)^n$$

$$= \left[q_1 + p_1 e^{t_1}\right]^n \quad \left(\because p_1 + q_1 = 1 \Rightarrow q_1 = 1 - p_1\right)$$

This is the <u>moment generating function</u> of Binomial distribution with parameters <u>n and p_1</u>.

Hence by uniqueness property of moment generating function, $X_1 \sim B(n, p_1)$.

Since each X_i is a binomial variate with parameters n and p_i so we have

$E(X_i) = np_i$

and $V(X_i) = np_i q_i \quad \forall\, i = 1, 2, \ldots, k$

CamScanner

$E(X_i) = np_i$

and $V(X_i) = np_i q_i \quad \forall\, i = 1, 2, \dots, k$

July 27, 2024

Characteristic function of multinomial distribution

If $X \sim M_K(n, p_1, p_2, \dots, p_K)$ then the pmf of the distribution is given by

$$P(X = x_1, x_2, \dots, x_K) = \frac{n!}{x_1!\, x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K}$$

$\phi_X(t) = E[e^{itX}]$

$E(X) = \sum_x x P(x)$

Now, the characteristic function of random variable X is given by

$$\phi_X(t) = \phi_{(x_1, x_2, \dots, x_K)}(t_1\, t_2 \dots t_K)$$

$$= E[e^{it_1x_1} e^{it_2x_2} \dots e^{it_Kx_K}]$$

$$= \sum_x [e^{it_1x_1} e^{it_2x_2} \dots e^{it_Kx_K}] \left[\frac{n!}{x_1!\, x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K}\right]$$

$$\Rightarrow \phi_X(t) = \sum_x \frac{n!}{x_1!\, x_2! \dots x_K!} (p_1 e^{it_1})^{x_1} (p_2 e^{it_2})^{x_2} \dots (p_K e^{it_K})^{x_K}$$

$$\boxed{\therefore \phi_X(t) = (p_1 e^{it_1} + p_2 e^{it_2} + \dots + p_K e^{it_K})^n}$$

$\sum_x \frac{n!}{x_1!\, x_2! \dots x_K!} p_1^{x_1} \dots p_K^{x_K} = (p_1 + p_2 + \dots + p_K)^n$

This is the characteristic function of multinomial distribution with parameters $(n, p_1, p_2, \dots, p_K)$.

Covariance and Correlation of multinomial distribution

* Covariance between X_i and X_j random variable in multinomial distribution

$Cov(X, Y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{n} = \frac{1}{n}\sum xy - \bar{x}\bar{y}$

$Cov(X, Y) = E(XY) - E(X)E(Y)$

<u>Solution</u> If $X \sim M_K(n, p_1, p_2, \dots, p_K)$ then the mgf of the distb^n is given by

$$M_X(t) = M_{X_1, X_2, \dots, X_K}(t_1\, t_2 \dots t_K)$$

$$= (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_K e^{t_K})^n$$

Now, the covariance between X_i and X_j, where $i \neq j = 1, 2, \dots, k$

$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$

Scanned with CamScanner

$$= E(X_i X_j) - (np_i)(np_j)$$

$E(XY)$
$= E(X,Y)$
$= E(X \text{ and } Y)$

Now, $E(X_i X_j) = \left.\dfrac{d^2 M_X(t)}{dt_i\, dt_j}\right|_{t=0}$

PARTIAL DERIVATIVE

$$\Rightarrow E(X_i X_j) = \left\{ \frac{d}{dt_i}\left[\frac{d\ M_X(t)}{dt_j}\right]\right\}_{t=0}$$

$$= \frac{d}{dt_i}\left[\frac{d}{dt_j}\left(p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_K e^{t_K}\right)^n\right]_{t=0}$$

$$= \frac{d}{dt_i}\left[n\left(p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_K e^{t_K}\right)^{n-1} \cdot p_j e^{t_j}\right]_{t=0}$$

$$\cancel{= np_j e^{t_j}\left[\frac{d}{dt_i}(n-1)\left(p_1 e^{t_1} + \dots + p_K e^{t_K}\right)^{n-2} \cdot p_i e^{t_i}\right.}$$

$$= np_j e^{t_j}\left[\frac{d}{dt_i}\left(p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_K e^{t_K}\right)^{n-1}\right]_{t=0}$$

$$= \left[(np_j e^{t_j})(n-1)\, p_i e^{t_i}\left(p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_K e^{t_K}\right)^{n-2}\right]_{t=0}$$

$$= (np_j)(n-1)(p_i)\left(p_1 + p_2 + \dots + p_K\right)^{n-2}$$

$$= n(n-1)\, p_i p_j$$

$$\therefore Cov(X,Y) = n(n-1)\, p_i p_j - (np_i)(np_j)$$

$$= \cancel{n^2 p_i p_j} - np_i p_j - \cancel{n^2 p_i p_j}$$

$$\boxed{\therefore Cov(X,Y) = -np_i p_j}$$

i.e. the multinomial distribution is <u>negatively correlated</u>.

Scanned with CamScanner

Correlation

The correlation between X_i and X_j

$$r(X_i, X_j) = \frac{Cov(X_i, X_j)}{sd(X_i) \cdot sd(X_j)}$$

$$= \frac{-n p_i p_j}{\sqrt{n p_i (1-p_i)} \sqrt{n p_j (1-p_j)}}$$

$$= -\left(\frac{p_i p_j}{(1-p_i)(1-p_j)}\right)^{1/2}$$

$E(X_i) = n p_i$

$V(X_i) = n p_i q_i$

$= n p_i (1-p_i)$

$sd(X_i) = \sqrt{n p_i (1-p_i)}$

Variance and covariance matrix of multinomial distribution

$$\begin{array}{c|cccc} & 1 & 2 & \cdots & k \\ \hline 1 & n p_1 q_1 & -n p_1 p_2 & \cdots & -n p_1 p_k \\ 2 & -n p_2 p_1 & n p_2 q_2 & \cdots & -n p_2 p_k \\ \vdots & \vdots & \vdots & & \vdots \\ k & -n p_k p_1 & -n p_k p_2 & & n p_k q_k \end{array}$$

Same, represents variance

Different, represents covariance

(O) — Variance

Remaining, Covariance

Problems

Q1. An urn contains 8 red balls, 3 yellow balls and 9 white balls. 6 balls are randomly selected with replacement. What is the probability that 2 are red, 1 is yellow and 3 are white?

Solution

Total number of balls = 20

No. of red balls = 8

No. of yellow balls = 3

Scanned with CamScanner

Number of white balls = 9

Since, 6 balls are randomly selected with replacement, so the number of trials, $n = 6$.

Let, p_1 = probability of getting red ball $= \frac{8}{20} = 0.4$

p_2 = probability of getting yellow ball $= \frac{3}{20} = 0.15$

p_3 = probability of getting white ball $= \frac{9}{20} = 0.45$

Here, $X_i \sim M(n, p_1, p_2, p_3)$ so, $P(X_1 = x_1, X_2 = x_2, \dots X_K = x_K) = \frac{n!}{x_1! \dots x_K!} p_1^{x_1} \cdots p_K^{x_K}$

Now, the probability of getting 2 Red, 1 yellow and 3 white balls,

$$P(X_1 = 2, X_2 = 1, X_3 = 3) = \frac{n!}{x_1!\, x_2!\, x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$= \frac{6!}{2!\,1!\,3!} (0.4)^2 (0.15)^1 (0.45)^3$$

$$= 60 \times 0.16 \times 0.15 \times 0.091125$$

$$= \boxed{0.13122}$$

Q A fair dice is thrown five times.

(i) What is the probability of getting one 6, one 5, two 4s, one 3 and no others?

(ii) What is the probability that the sum of the scores is exactly 8?

Solution

(i) p_1 = probability of getting 6 = 1/6

p_2 = " " " 5 = 1/6

p_3 = " " " 4 = 1/6

p_4 = " " " 3 = 1/6

no. of trials, $n = 5$

Here, $X_i \sim M(n, p_1, p_2, p_3, p_4)$ so, $P(x_1, x_2, \dots x_K)$

$$= \frac{n!}{x_1! \dots x_K!} p_1^{x_1} \dots p_K^{x_K}$$

Now, probability of getting one 6, one 5, two 4s and one 3, $P(X_1 = 1, X_2 = 1, X_3 = 2, X_4 = 1) = \frac{n!}{x_1!\, x_2!\, x_3!\, x_4!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$

$$= \frac{5!}{1!\,1!\,2!\,1!} (1/6)^1 (1/6)^1 (1/6)^2 * (1/6)^1$$

$$= 60 * (1/6)^5$$

$$= \boxed{0.007716}$$

(ii)

Sum 8 can be obtained either of the following ways:

→ 1, 1, 1, 1, 4.
$x_1 = 4, x_4 = 1$ or

→ 1, 1, 1, 2, 3
$x_1 = 2, x_2 = 1, x_3 = 1$ OK

→ 1, 1, 2, 2, 2
$x_1 = 2, x_2 = 3$

Scanned with CamScanner

PRIOR AND POSTERIOR DISTRIBUTION

Multiplication law of probability

Let, A and B be two dependent events then, the probability of getting events A and B is given by

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B/A)$$

INITIALLY OCCURS

$$\text{or,} \quad P(A \cap B) = P(B) \cdot P(A/B)$$

where, $P(A/B)$ = Conditional probability of A given B

$P(B/A)$ = Conditional probability of B given A.

If events A and B are independent, then,

$$P(A \cap B) = P(A) \cdot P(B)$$

$P(A \cap B) = P(A) \cdot P(B/A)$

$P(B/A) = \dfrac{P(A \cap B)}{P(A)}$; $P(A) \neq 0$

Conditional Probability

Conditional probability refers to the happening of an event when the information about happening of other event is given.

Let, A and B be two dependent events, then the conditional probability of A given B,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \; ; \; P(B) \neq 0$$

Similarly, $P(B/A) = \dfrac{P(A \cap B)}{P(A)}$; $P(A) \neq 0$

If A and B are independent then,

$P(A/B) = P(A)$ and $P(B/A) = P(B)$ $[\because P(A \cap B) = P(A) \cdot P(B)]$

- QUESTIONS -

Q1.

Let, A = event that accidents are alcohol related
B = event that accidents occured at night

$P(B) = 0.6$
$P(A) = 0.52$
$P(A \cap B) = 0.37$

a. Probability that an accident was alcohol related given that it occured at night,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.37}{0.6} = \underline{0.616}$$

i.e. if accidents occured at night then there is 61% chances that the accidents are alcohol related.

b.

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.37}{0.52} = \underline{0.711}$$

Q2.

Let, A = event in which pilots strike
B = event in which drivers strike

$P(A) = 0.75$
$P(B) = 0.65$
$P(A/B) = 0.90$

a. $P(A \cap B) = P(B) \cdot P(A/B)$

$= 0.65 * 0.90$

$= \underline{0.585}$

b. $P(B/A) = \frac{P(A \cap B)}{P(A)}$

$= \frac{0.585}{0.75}$

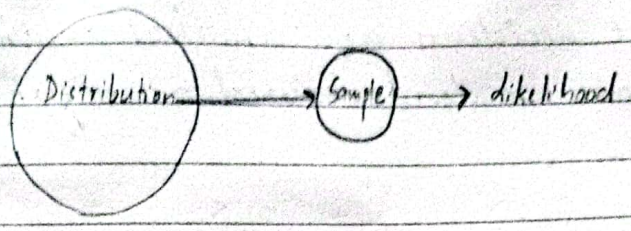
$= \underline{0.78}$

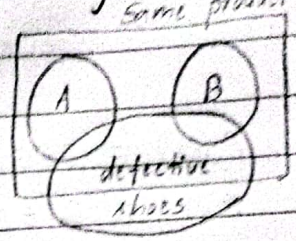
Scanned with CamScanner

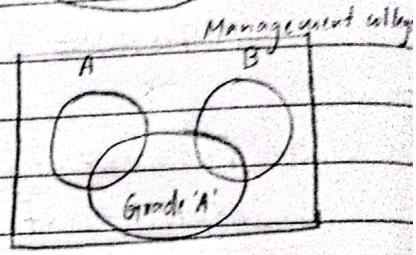
Q3.

Common event

Mutually exclusive event







Let, E_1 = screws manufactured by plant I

E_2 = screws manufactured by plant II.

S = Standard quality screws manufactured by plant

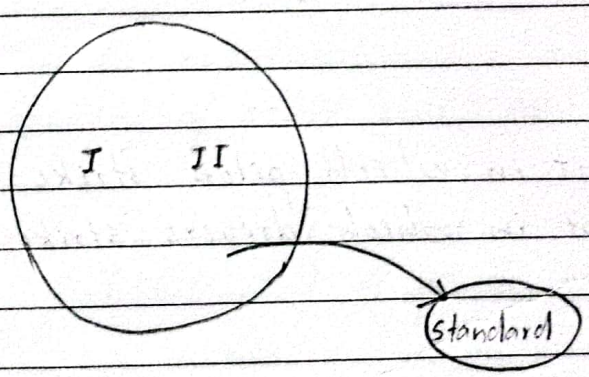
$S \subset E_1 \cup E_2$

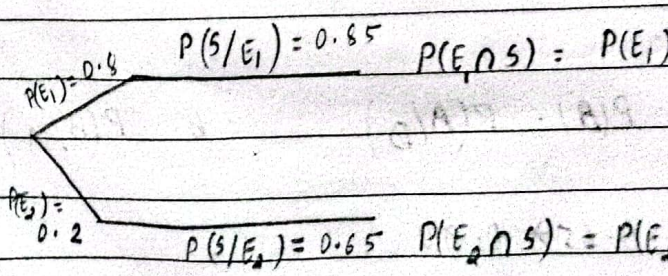
Given, $P(E_1) = 0.80$

$P(E_2) = 0.20$

$P(S/E_1) = 0.85$

$P(S/E_2) = 0.65$



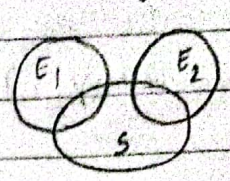


$P(E_1 \cap S) = P(E_1) \cdot P(S/E_1) = 0.8 * 0.85 = 0.68$

$P(E_2 \cap S) = P(E_2) \cdot P(S/E_2) = 0.2 * 0.65 = 0.13$

Probability of getting standard quality screw is given by

$$P(S) = P(E_1 \cap S) + P(E_2 \cap S)$$
$$= 0.68 + 0.13$$
$$= 0.81$$



(a) Probability that a randomly selected standard quality screw is manufactured by plant I,

$$P(E_1/S) = \frac{P(E_1 \cap S)}{P(S)}$$
$$= \frac{0.68}{0.81}$$
$$= \underline{0.839}$$

(b) ... manufactured by plant II

$$P(E_2/S) = \frac{P(E_2 \cap S)}{P(S)}$$
$$= \frac{0.13}{0.81}$$
$$= \underline{0.1604}$$

or $1 - P(E_1/S)$

$$= 1 - 0.839$$
$$= 0.16$$

$$\boxed{P(A/B) + P(A'/B) = 1}$$

Q11. H = Husband is watching TV
H' = Husband is not watching TV
W = Wife is watching TV

$P(H) = 0.60$, $P(H') = 0.40$

Q12.

a. G = Event that football team will play on grass
A = " " " " " " " artificial turf
K = Event that the player incurs a knee injury

given, $P(A) = 0.3 \Rightarrow P(G) = 0.70$

$P(K/G) = 0.40$

$P(K/A) = 0.40 + 60\%$ of 0.40

$= 0.40 + 0.24$

$= 0.64$

a. The probability that a randomly selected player incurs a knee injury,

$P(K) = P(G \cap K) + P(A \cap K)$

$= P(G) \cdot P(K/G) + P(A) \cdot P(K/A)$

$= 0.70 * 0.40 + 0.30 * 0.64$

$= \underline{0.472}$

b. The probability that a randomly selected football player with a knee injury incurred the injury playing on grass,

$$P(G/K) = \frac{P(G \cap K)}{P(K)}$$

$$= \frac{P(G) \cdot P(K/G)}{P(K)} = \frac{0.7 * 0.4}{0.472}$$

$= 0.593$

Q13. Let, W = Women with breast cancer

W' = Women without breast cancer

T = Test is positive

Common event $\longrightarrow$ Mutually exclusive events

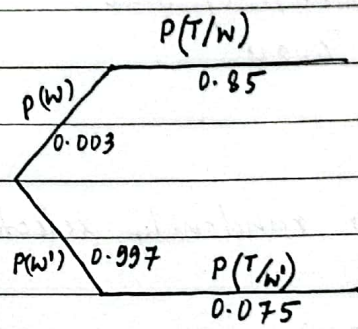
Given,

$P(T/W) = 0.85$, $P(W) = 0.3\% = 0.003$

$P(W') = 1 - 0.003 = 0.997$

$P(T'/W') = 0.925 \Rightarrow P(T/W') = 1 - 0.925$ $\left[\because P(A/B) + P(A'/B) = 1\right.$

$= 0.075$



$P(W \cap T) = 0.003 * 0.85 = 0.00255$

$P(W' \cap T) = 0.997 * 0.075 = 0.074$

Probability that the test is positive $P(T) = P(W \cap T) + P(W' \cap T)$

$= 0.00255 + 0.074$

$= 0.0765$

If the test is positive, then the probability that the woman has cancer is given by

$$P(W/T) = \frac{P(W \cap T)}{P(T)}$$

$$= \frac{P(W) \cdot P(T/W)}{P(T)}$$

$$= \frac{0.003 * 0.85}{0.0765}$$

$$= \frac{0.0025}{0.0765}$$

$$= \underline{0.032}$$

Statement of Baye's theorem

Let $E_1, E_2, \ldots, E_n$ be mutually exclusive events defined on the sample space 's' such that $P(E_i) > 0 \ \forall i = 1, 2, \ldots$ Let, A be any arbitrary event which is a subset of $\bigcup_{i=1}^{n} E_i$ such that $P(A) > 0$ then Baye's theorem states that,

$$P(E_i/A) = \frac{P(E_i \cap A)}{P(A)}$$

$$= \frac{P(E_i) \cdot P(A/E_i)}{P(A)}$$

where,

$$P(A) = P(A \cap E_1) + \cdots + P(A \cap E_n)$$

$$= P(E_1) \cdot P(A/E_1) + \cdots + P(E_n) \cdot P(A/E_n)$$

$$= \sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)$$

Likelihood Function

The joint probability mass function or joint probability density function is called likelihood function.

Let $X = x_1, x_2, \ldots, x_n$ be a sample drawn from a population with probability density function $f(x; \theta)$, then the likelihood function of sample observations $x_1, x_2, \ldots, x_n$ is given by

$$L = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta)$$

$$= \prod_{i=1}^{n} f(x_i; \theta)$$

[Here, x_i's are identically, independently distributed (iid)].

Example:

\# Let $x_i \sim P(\lambda)\ \forall\ i = 1, 2, \dots, n$, then, the likelihood function of $x_1, x_2, \dots, x_n$ is given by

$$L = \prod_{i=1}^{n} P(x_i ; \lambda)$$

$$= P(x_1 ; \lambda) \dots P(x_n ; \lambda)$$

$$= \frac{e^{-\lambda} \lambda^{x_1}}{x_1 !} \dots \frac{e^{-\lambda} \lambda^{x_n}}{x_n !}$$

$$= \frac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\prod_{i=1}^{n} (x_i !)}$$

\# Let $x_i \sim N(\mu, \sigma^2)$ ← Normal distribution with parameter μ and σ^2.

then, the likelihood function of $x_1, x_2, \dots, x_n$ is given by:

$$L = \prod_{i=1}^{n} f(x_i ; \mu, \sigma^2)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left(\frac{x_1 - \mu}{\sigma}\right)^2} \dots \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left(\frac{x_n - \mu}{\sigma}\right)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2}$$

If $X \sim N(\mu, \sigma^2)$ then

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left(\frac{x - \mu}{\sigma}\right)}$$

; $-\infty < x < \infty$

$-\infty < \mu < \infty$

$\sigma > 0$

- PRIOR AND POSTERIOR DISTRIBUTION -

In Bayesian Inference, we treat the parameter 'θ' as a random variable having probability density function or probability mass function $f(\theta)$; $\theta \in \Theta$. The probability distribution of

Statistics

Descriptive | Inference

Decision Making

'θ' defines the prior probability density/mass function $f(\theta)$ is called prior distribution

Let us consider a random sample $X = x_1, x_2, \ldots, x_n$ be drawn from a population with pdf/pmf $f_\theta(x)$ where θ is considered as a parameter θ.

Aug 3, 2024

NOTE: If $X \sim B(m, n)$ then

$$f(x) = \frac{x^{m-1}(1-x)^{n-1}}{\beta(m, n)}$$

If $X \sim G(\alpha, \beta)$, then

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma\alpha} \quad ; \quad x > 0$$

If $X \sim \exp(\theta)$

$$f(x) = \theta e^{-\theta x} \quad ; \quad x > 0$$

$$\beta(m, n) = \frac{\Gamma m \, \Gamma n}{\Gamma m+n}$$

$$\Gamma m = (m-1)!$$

If $X \sim Gpo(\theta)$

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

$$E(X) = \frac{1}{\theta}$$

If $X \sim Gpo(\theta)$

$$f(x) = \frac{1}{\theta} e^{-1/\theta x}$$

$$E(X) = \theta$$

(Continue)

Here, we denote the pdf/pmf of X, $f_\theta(x)$ by $f(x/\theta)$, the conditional pdf/pmf of X given θ, called likelihood

function of x/θ

Now, the pdf/pmf of x and θ is given by

$$f(x, \theta) = f(\theta) \cdot f(x/\theta)$$

$\therefore$ The marginal pdf/pmf of x given by

$$f(x) = \int_\theta f(x, \theta)\, d\theta$$

$$= \int_\theta f(\theta) \cdot f(x/\theta)\, d\theta$$

[For continuous]

wt. → X \ height → Y	y_1	y_2	...	y_m
x_1	P_{11}	P_{12}	...	P_{1m}
⋮				
x_n				

$$P(X = x_1) = \sum_{j=1}^{m} P_{1j}$$

$$f(x) = \sum_\theta f(\theta)\, f(x/\theta) \quad \text{[For discrete]}$$

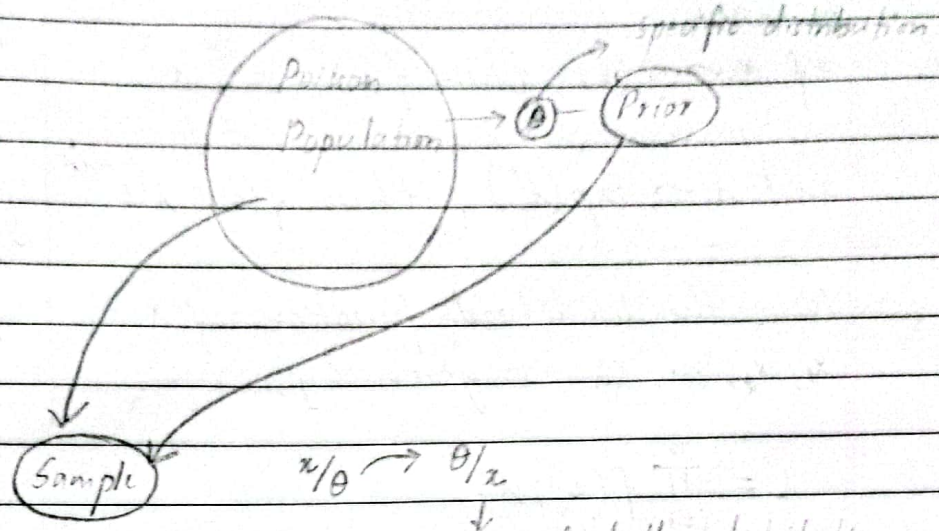
Now, applying Bayes theorem, we get

$$f(\theta/x) = \frac{f(x, \theta)}{f(x)}$$

$$\boxed{\therefore f(\theta/x) = \frac{f(\theta) \cdot \overset{f(x/\theta)}{f(x/\theta)}}{f(x)}}$$

The conditional probability density function $f(\theta/x)$ is called posterior probability density function and the distribution defined by posterior probability function $f(\theta/x)$ is called posterior distribution.

The prior distribution gives the distribution of θ before the sample is drawn while the posterior distribution of θ gives the conditional distribution of θ after the sample is drawn, so the posterior distrib^n reflects both the prior information and the sample information.



$$f(\theta/x) \propto \underbrace{f(\theta)}_{\text{Prior}} \cdot \underbrace{f(x/\theta)}_{\text{Likelihood}}$$

NOTE: A usual way of expressing the posterior density function is to use proportionality.

We have,

$$\underbrace{f(\theta/x)}_{\text{Posterior of } \theta} = \frac{\overbrace{f(\theta)}^{\text{Prior}} \cdot \overbrace{f(x/\theta)}^{\text{Likelihood}}}{f(x) \rightarrow \text{Normalizing constant}}$$

$$\Rightarrow f(\theta/x) \propto f(\theta) \cdot f(x/\theta)$$

$$\Rightarrow \underbrace{f(\theta/x)}_{\text{Posterior}} \propto \underbrace{f(\theta)}_{\text{Prior}} * \text{Likelihood}$$

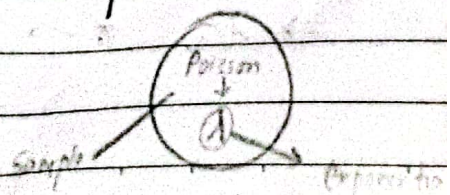
$\Rightarrow$ Posterior $\propto$ Prior * likelihood.

EXAMPLE

Let $X/\lambda \sim P(\lambda)$ and prior distribution of 'λ' is exponential with parameter 'λ'.

Find the posterior distribution of λ.

$[f(\lambda/x) = ?]$



Solution

If $X \sim P(\lambda)$, then $f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

and $\lambda \sim Gp(\lambda')$, then $f(\lambda) = \lambda' e^{-\lambda'\lambda}$

If $X \sim P(\lambda)$,
$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

If $X \sim Gp(\theta)$
$f(x) = \theta e^{-\theta x}$

Now, the likelihood function of $X = x_1, x_2, \dots, x_n$ is given by,

$$L = \prod_{i=1}^{n} f(x_i; \lambda)$$

$$= f(x_1; \lambda) \dots f(x_n; \lambda)$$

$$\Rightarrow L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \dots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i)!} \rightarrow \text{Constant}$$

$$L = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i} \cdot \text{constant}$$

Now, the prior pdf of λ is

$$f(\lambda) = \lambda' e^{-\lambda'\lambda}$$

$$= e^{-\lambda'\lambda} \cdot \text{constant}$$

Now, the posterior distribution of λ is calculated as

Posterior $\propto$ Prior * likelihood

$$\Rightarrow f(\lambda/x) \propto \{e^{-\lambda'\lambda}\} * \{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}\}$$

$$\Rightarrow f(\lambda/x) \propto e^{-\lambda(n+\lambda')}\lambda^{\left(\sum_{i=1}^{n} x_i + 1\right)-1}$$

$$\Rightarrow \lambda/x \sim G\left(\sum_{i=1}^{n} x_i + 1, n + \lambda'\right)$$

If $X \sim G(\alpha, \beta)$ then,
$f(x) = \frac{e^{-\beta x} x^{\alpha-1} \beta^{\alpha}}{\Gamma\alpha}$

Q. The annual number of claims arising from a particular group of policies follow a Poisson distribution with mean μ. The prior distribution of μ is exponential with mean 30.
In the previous two years, the number of claims arising from the group were 28 and 26 respectively. Determine the posterior distribution of μ.

Solution

X = annual number of claims

Here, X follows Poisson distribution i.e. $X \sim P(\mu)$.

so, $P(X = x) = \dfrac{e^{-\mu}\mu^{x}}{x!}$

Now, the likelihood function,

$$L = \prod_{i=1}^{2} f(x_i, \mu)$$

$$= f(x_1, \mu) * f(x_2, \mu)$$

$$= \frac{e^{-\mu}\mu^{x_1}}{x_1!} * \frac{e^{-\mu}\mu^{x_2}}{x_2!}$$

$$= e^{-2\mu} \cdot \mu^{x_1 + x_2} \cdot \text{constant} \quad [\because x_1 = 28, x_2 = 26]$$

$$\Rightarrow L = e^{-2\mu}\mu^{54} \cdot \text{constant}$$

$\mu \sim$ Exp with mean 30

$f(\mu) = \dfrac{1}{30} e^{-\mu/30}$

Further, μ follows exponential distribution with mean 30.

$\therefore f(\mu) = \dfrac{1}{30} e^{-\mu/30}$

$$= e^{-\mu/30} \cdot \text{constant}$$

Now, posterior distribution of μ is calculated as

Posterior $\propto$ Prior $*$ likelihood

$$\Rightarrow f(\mu/x) \propto \{e^{-\mu/30}\} * \{e^{-24\mu} \mu^{54}\}$$

$$\Rightarrow f(\mu/x) \propto e^{-61/30\,\mu} \mu^{55-1}$$

$\therefore$ μ/x follows gamma distribution with $\alpha = 55$, $\beta = \frac{.61}{30}$

$$\therefore \mu/x \sim G(55, .61/30)$$

August 6, 2021

Q. If $X/\lambda \sim Exp(\lambda)$ and prior of $\lambda \sim Exp(\lambda')$ find the posterior distribution of λ.

Solution Since, $X \sim Exp(\lambda)$ so the likelihood function L of sample observations $x_1, x_2, \ldots, x_n$ is	If $X \sim Exp(\theta)$ $f(x) = \theta e^{-\theta x}$ (Posterior $\propto$ Prior $\times$ likelihood

$$L = \prod_{i=1}^{n} f(x_i; \lambda) = \{\lambda e^{-\lambda x_1}\} \ldots \{\lambda e^{-\lambda x_n}\}$$

$$= \lambda^n e^{-\lambda \sum x_i}$$

Further,

$$f(\lambda) = \lambda' e^{-\lambda' \lambda}$$

$$= e^{-\lambda' \lambda} \cdot \text{constant}$$

Now, the posterior distribution of λ is obtained as

Posterior $\propto$ Prior $\times$ likelihood

$$\Rightarrow f(\lambda/x) \propto \{e^{-\lambda'\lambda}\}\{\lambda^{n} e^{-\lambda \sum_{i=1}^{n} x_i}\}$$

$$\Rightarrow f(\lambda/x) \propto e^{-\lambda(\lambda' + \sum x_i)} \lambda^{(n+1)-1}$$

$$\Rightarrow \lambda/x \sim G(n+1, \lambda' + \sum x_i)$$

Assignment

If $X/\lambda \sim G(\alpha, \lambda)$ and prior $\lambda \sim G(\alpha', \lambda')$ then find the posterior distribution of λ.

Scanned with CamScanner

Q. If $X/p \sim Bin(m, p)$ and prior $p \sim \beta(\alpha, \lambda)$, then find the posterior distribution of 'p'.

Solution

Since $X/p \sim B(m, p)$ so

$$f(X/p) = {}^{m}C_{x}\, p^{x} (1-p)^{m-x}$$

If $X \sim B(n, p)$

then $f(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$

$X \sim \beta(\alpha, \lambda)$

$$f(x) = \frac{x^{\alpha-1} (1-x)^{\lambda-1}}{\beta(\alpha, \lambda)}$$

$$\beta(m, n) = \frac{\Gamma m \, \Gamma n}{\Gamma m+n}, \quad \Gamma m = (m-1)!$$

Now, the likelihood function L of sample observations $x_1, x_2, \ldots, x_n$

$$L = \prod_{i=1}^{n} f(x_i; m, p)$$

$$= \left\{ {}^{m}C_{x_1}\, p^{x_1} (1-p)^{m-x_1} \right\} \cdots \left\{ {}^{m}C_{x_n}\, p^{x_n} (1-p)^{m-x_n} \right\}$$

$$= \left\{ p^{\sum_{i=1}^{n} x_i} \right\} * \left\{ (1-p)^{mn - \sum_{i=1}^{n} x_i} \right\} \cdot \text{constant}$$

Further, $p \sim \beta(\alpha, \lambda)$

$$\therefore f(p) = \frac{p^{\alpha-1} (1-p)^{\lambda-1}}{\beta(\alpha, \lambda)} = p^{\alpha-1} \cdot (1-p)^{\lambda-1} \cdot \text{constant}$$

Now, the posterior distribution of p is calculated as

Posterior $\propto$ Prior * likelihood

$$\Rightarrow f(p/x) \propto \left\{ p^{\alpha-1} \cdot (1-p)^{\lambda-1} \right\} \left\{ p^{\sum_{i=1}^{n} x_i} (1-p)^{mn - \sum_{i=1}^{n} x_i} \right\}$$

$$\Rightarrow f(p/x) \propto p^{\left(\sum_{i=1}^{n} x_i + \alpha\right) - 1} (1-p)^{\left(mn - \sum_{i=1}^{n} x_i + \lambda\right) - 1}$$

$$\Rightarrow p/x \sim \beta\left(\alpha + \sum_{i=1}^{n} x_i, \; mn - \sum_{i=1}^{n} x_i + \lambda\right)$$

Hence, posterior of 'p' follows Beta distribution with parameters $\left(\alpha + \sum_{i=1}^{n} x_i\right)$ & $\left(mn - \sum_{i=1}^{n} x_i + \lambda\right)$.

CONJUGATE PRIOR

For a given likelihood, if the posterior distribution belongs to the same family as the prior distribution then prior is called conjugate prior.

Q. The no. of claims per month has a negative binomial distribution with parameters k and p. The number of claims observed over 'n' months are $x_1, x_2, \ldots, x_n$.

(i) Show that the beta distribution is the conjugate prior for 'p'.

If $X \sim NB(k,p)$
$f(x) = {}^{x+k-1}C_x \, p^k (1-p)^x$

If $X \sim \beta(m,n)$, then
$f(x) = \dfrac{x^{m-1}(1-x)^{n-1}}{\beta(m,n)}$

Solution

Given, $X \sim NB(k,p)$ so,

$$f(x) = {}^{x+k-1}C_x \, p^k (1-p)^x$$

Now, the likelihood function

$$L = \prod_{i=1}^{n} f(x_i; k, p)$$

$$= \left\{ {}^{x_1+k-1}C_{x_1} \, p^k (1-p)^{x_1} \right\} \cdots \left\{ {}^{x_n+k-1}C_{x_n} \, p^k (1-p)^{x_n} \right\}$$

$$= \left\{ p^{nk} \right\} \left\{ (1-p)^{\sum_{i=1}^{n} x_i} \right\} \cdot \text{constant}$$

Further, $p \sim \beta(\alpha, \lambda)$

$$\therefore f(p) = \frac{p^{\alpha-1}(1-p)^{\lambda-1}}{\beta(\alpha,\lambda)}$$

$$= p^{\alpha-1}(1-p)^{\lambda-1} \cdot \text{constant}$$

Now, the posterior of p is calculated as

Posterior $\propto$ Prior $*$ Likelihood

$$\Rightarrow f(p/x) \propto \left\{ p^{\alpha-1} \cdot (1-p)^{\lambda-1} \right\} \left\{ p^{nk} (1-p)^{\sum_{i=1}^{n} x_i} \right\}$$

$$\Rightarrow f(p/x) \propto p^{(\alpha + nk) - 1} (1-p)^{(\sum_{i=1}^{n} x_i + \lambda) - 1}$$

$$\Rightarrow p/x \sim \beta\left(\alpha + nk, \sum_{i=1}^{n} x_i + \lambda\right)$$

Hence, posterior of p follows beta distribution. Thus, beta distribution is the conjugate prior for p.

Assignment

Q. 10 iid observations from a $P(\lambda)$ distribution gave 3, 4, 3, 1, 5, 5, 2, 3, 3, 2. Assuming an $Exp(0.2)$ as prior distribution of λ.
Find the posterior distribution of λ.

Q. A random sample of size 10 from a Poisson distribution with mean λ yields the following data values:
3, 4, 3, 1, 5, 5, 2, 3, 3, 2.
The prior distribution of λ is Gamma (5, 2).
Find the posterior distribution of λ and also find the mean of distribution.

$X \sim G(\alpha, \beta)$

$E(X) = \alpha/\beta$

August 8, 2024

Q. The number of claims in a week arising from a certain group insurance policies has a Poisson distⁿ with mean λ. Seven claims were incurred in the last week. The prior distⁿ of λ is uniform on the integers 8, 10, and 12.

(i) Determine the posterior distⁿ of λ.

(ii) Find the mean of posterior distⁿ.

Solution

Let X = number of claims in a week

Here $X \sim P(\lambda)$ so $P(X = x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

Further, λ is uniform on the integers 8, 10, & 12.

$\therefore$

λ:	8	10	12
$P(\lambda)$	1/3	1/3	1/3

To obtain the posterior distribution, we must calculate $P(\lambda = 8/X = 7)$, $P(\lambda = 10/X = 7)$ and $P(\lambda = 12/X = 7)$

Now, probability of

$P(\lambda = 8/X = 7) \propto P(\lambda = 8) \cdot P(X = 7/\lambda = 8)$

$$= \frac{1}{3} \cdot \left(\frac{e^{-8} 8^{7}}{7!}\right)$$

$$= \underline{0.0465}$$

$P(\lambda = 10/X = 7) \propto P(\lambda = 10) \cdot P(X = 7/\lambda = 10)$

$$= \frac{1}{3} \cdot \left(\frac{e^{-10} 10^{7}}{7!}\right) = \underline{0.0300}$$

$P(\lambda = 12 / x = 7) \propto P(\lambda = 12) \cdot P(x = 7 / \lambda = 12)$

$$= \frac{1}{3} \cdot \left(\frac{e^{-12} \, 12^{7}}{7!} \right)$$

$$= \underline{0.014}$$

$\therefore P(x = 7) = P(x = 7, \lambda = 8) + P(x = 7, \lambda = 10) + P(x = 7, \lambda = 12)$

$= 0.465 + 0.03 + 0.014$

$= \cancel{0.509} \quad \underline{0.091}$

Now, the posterior distⁿ

$P(\lambda = 8 / x = 7) = \dfrac{0.0465}{0.091} = 0.510$

$P(\lambda = 10 / x = 7) = \dfrac{0.030}{0.091} = 0.329$

$P(\lambda = 12 / x = 7) = \dfrac{0.014}{0.091} = 0.153$

(i)

λ	8	10	12
λ / x	0.510	0.329	0.153

(ii) Now, expected value of posterior distⁿ

$E(\lambda / x) = \Sigma \lambda \, P(\lambda / x)$

$= 8(0.510) + 10(0.329) + 12(0.153)$

$= \underline{9.206}$

Assignment

The no. of claims received per day from a certain portfolio has a Poisson distⁿ with mean λ. The prior distⁿ of λ is as follows:

λ :	1	2	3
$P(\lambda)$:	0.3	0.5	0.2

Given that 3 claims were received last day, determine the posterior distⁿ of λ and the mean.

Hypothesis

↓

Assumption for population parameter

Avg wt: $\boxed{55\ kg}$

Sample

Statistic

	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Proportion	P	p
Correlation	ρ	r

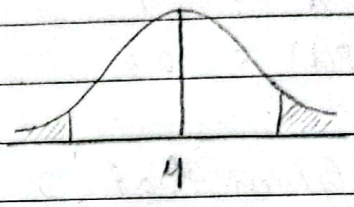
Types of Hypothesis

i) Null Hypothesis

$H_0 : \mu = \mu_0$

ii) Alternative Hypothesis

$H_1 : \mu \neq \mu_0$ (Two tailed test)



$H_1 : \mu > \mu_0$ (Right tailed test) → One tailed test

$H_1 : \mu < \mu_0$ (Left tailed test)

eg: 55 kg or not

eg: ~~more~~ Less than 55 kg

eg: more than 55

Types of Error

Situation / Decision making	H_0 is true	H_0 is false
Reject H_0	Wrong decision (Type I error)	Correct decision
Do not reject H_0	Correct decision	Wrong decision (Type II error

* Type first error

α = Prob. (type first error) = Level of significan

also called <u>producer's risk.</u>

* Type second error

β = Prob (type second error)

Consumer's risk.

β = Prob (accept H_0 / H_0 is false)

$1-\beta$ = Prob (reject H_0 / H_0 is false)

= Prob (correct decision)

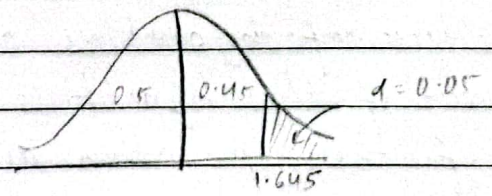
= Power of the test.

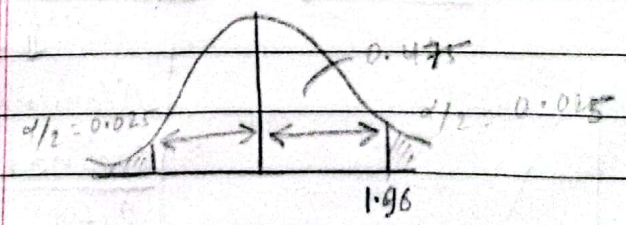
Critical value (significant value)

The value which separates the acceptance & rejection region is called the critical value.

Depends on

* Size of α (level of significance)
* Whether the test is one tailed or two tailed.





Degree of freedom

Number of values in the sample that can be choosen freely. For 'n' sample observations, the degree of freedom is (n-1).

↓

For small sample sized test

Identification of one tailed and two tailed test.

Aug 10, 2024

UNIT 6 : Non-parametric test: (NP-Test)

Parametric test

t - test
Z - test
F - test

} Follows normal distribution

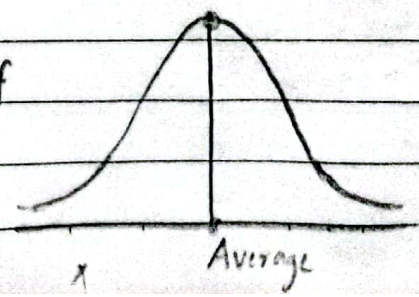
* Introduction

Non-parametric tests are defined as those statistical test of hypothesis in which parametric values are not involved.

In other words, those tests of hypothesis are said to be non-parametric test if the hypothesis does not involve any parameter of the population and if measurements are on the nominal or ordinal scale.

As $n \to \infty$,

Z



Scaling ⇓

Nominal, Ordinal	Qualitative data
Ratio, Interval	Quantitative data

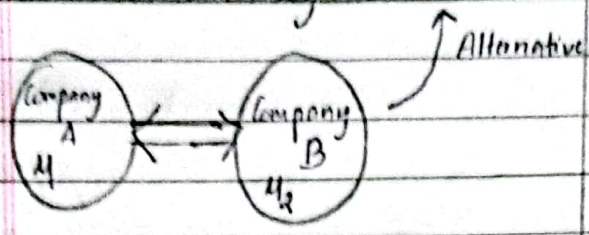
(VVI)

Difference between parametric and non-parametric test

Parametric	Non-parametric
i) Parametric tests are those tests whose models specify certain conditions about the parameter of the population from which the sample has been drawn.	i) N-P tests are those tests whose models do not specify any conditions about the parameters of the population from which the sample has been drawn.
(ii) Parametric tests are mostly applied only to the data which are measured in interval or ratio scale.	ii) N-P tests are applied only to data which are measured in nominal and ordinal scale.
(iii) Parametric tests are the most powerful.	(iii) N-P tests are less powerful or weaker than the parametric tests.
iv) Parametric tests are designed to test the statistical hypothesis of one or more parameters of the population.	iv) N-P tests are designed only to test statistical hypothesis which does not involve any parameter of the population.

Scanned with CamScanner

Mann-Whitney U test



Small sample size	Large sample size
n_1 and $n_2 \leq 10$	$n_1 > 10$, $n_2 > 10$
$n_1 + n_2 \leq 20$	$n_1 + n_2 > 20$
	z-test

↳ $Md_1 = Md_2$ ✓

or, $E(X) = E(Y)$

This test is the most powerful non-parametric test for testing hypothesis of difference between two independent location of two independent random samples. This

average

is non-parametric alternative test of t-test for difference of means.

Procedure:

Null hypothesis (H_0): $Md_1 = Md_2$ i.e. there is no significant difference between two sample medians or two pop^n medians.

Alternative hypothesis (H_1): $Md_1 \neq Md_2$ i.e. there is a significant difference between two sample medians or two population medians. (Two tailed test)

or $H_1: Md_1 > Md_2$ (Right) } one tailed test

or, $H_1: Md_1 < Md_2$ (Left)

Sample size ($n_1 \leq 10$ & $n_2 \leq 10$)

Test-statistic: Under H_0,

U_0 = minimum of $\{U_1$ and $U_2\}$

where $U_1 + U_2 = n_1 n_2$

p-value $\Rightarrow U_0, n_1, n_2$

x_1	x_2
10 (4)	15 (8.5)
18 (11)	12 (6.5)
12 (6.5)	19 (12)
15 (8.5)	25 (13)
17 (10)	6 (2)
5 (1)	9 (3)
	11 (5)
$n_1 = 6$	$n_2 = 7$

for 12, $\frac{6+7}{2} = 6.5$

Common rank

(Take average)

Increasing or ascending order

Rank of x_1 Rank of x_2

Sum of ranks, $R_1 =$, $R_2 =$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

→ where, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

R_1 = sum of ranks given to first sample data.

R_2 = " " " " " second " ".

P-value > α, then we do not reject H_0 or insignificant

P-value $\leq \alpha$ then we reject H_0 or significant

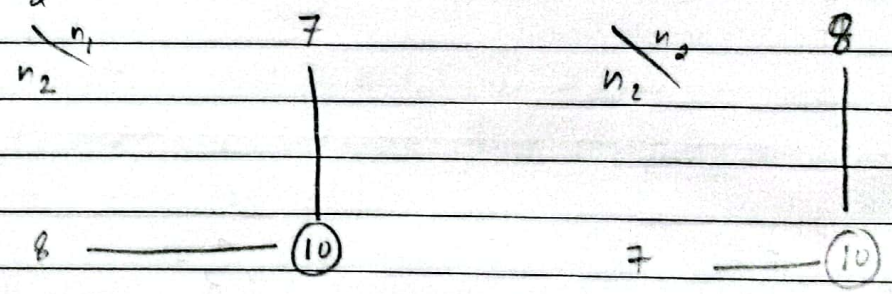
Critical value: The critical value of the test-statistic is obtained from Mann-Whitney U-test table at α level of significance and $(n_1 \& n_2)$.

Decision: If $U_0 > U_{\alpha,(n_1, n_2)}$ then we do not reject H_0 otherwise reject H_0.

$\alpha = 0.05$

Two tailed

$n_1 = 7, \quad n_2 = 8$



p-value

$n_1 = 5 \qquad U_0 = 6$

$n_2 = 10$

$\alpha = 0.05$

p-value = 0.0097

Multiply by 2 for two-tailed test.

$2P = \alpha$

The p-value is obtained from Mann-Whitney probability table for (n_1, n_2) and U_0.

For two tailed, P-value = $2P_0$
For one tailed, P-value = P_0

If p-value > α, then we do not reject H_0 otherwise reject H_0.

August 13, 2024 [SMALL SAMPLE SIZE]

Q14.

$n_1 \leq 10$ & $n_2 \leq 10$

$n_1 = 5$
$n_2 = 4$
$\alpha = 0.05$ = level of significance

If level of significance 'α' is not given, then consider 5%.

H_0: $Md_1 = Md_2$ i.e. There is no difference between two average numbers of correction of trained and untrained managers.
H_1: $Md_1 \neq Md_2$ i.e. there is difference between two average numbers of correction of trained and untrained managers (Two tailed test)

Test statistic: Under H_0,
U_0 = minimum of $\{U_1$ and $U_2\}$

where, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$ (Note that, $n_1 n_2 = U_1 + U_2$)

Calculation:

Trained	Untrained	Ranks of trained (combined)	Ranks of untrained
71	110	7	9
64	70	4	5
75	53	6	3
45	57	1	2
92		8	
		$R_1 = 26$	$R_2 = 19$

Now,

$$U_1 = 5\times4 + \frac{5(5+1)}{2} - 26$$

$$= 20 + 15 - 26$$

$$= 9$$

$$U_2 = 5\times4 + \frac{4\times5}{2} - 19$$

$$= 20 + 10 - 19$$

$$= 11$$

$\therefore U_0 = 9$

<u>Critical value</u>: We have, $\alpha = 0.05$ and $(n_1, n_2) = (5,4)$

$\therefore U_{(5,4),\,0.05}$ (two-tailed test) $= 1$

<u>Decision</u>: Since $U_0 > U_{(5,4),\,0.05}$, so we do not reject H_0. Hence, there is no difference between the average number of correction of trained and untrained managers.

p-value = 0.4524

$\therefore$ p-value = 2(0.4524)

= 0.9048

$\alpha = 0.05$

$\therefore$ p-value > α, so we do not reject H_0.

Assignment

Q15 and 17

Q. The heart beating rate of 5 vegetarians and 5 non vegetarians are recorded below:

Vegetarians :	56	67	82	60	75
Non-vegetarians:	53	42	75	58	65

Is the mean heart beating rate of non vegetarians significantly high. Use Mann Whitney U test at $\alpha = 5\%$.

Solution. $n_1 = 5$, $n_2 = 5$

$\alpha = 0.05$

H_0: $Md_1 = Md_2$ i.e. the mean heart beating rate of non-vegetarian is not significantly high than that of vegetarian.

H_1: $Md_1 < Md_2$ i.e. the mean heart beating rate of non-vegetarian is significantly higher than vegetarian (one tailed test)

Test statistic : Under H_0

U_0 = minimum of $\{U_1, U_2\}$

where, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

Calculation

Vegetarian	Non-vegetarian	Ranks of vegetarian (Combined)	Ranks of non
56	53	3	
67	42	7	
82	75	10	
60	58	5	
75	65	8.5	
		$R_1 = 33.5$	

Now,

$$U_1 = 5\times5 + \frac{5\times6}{2} - 33.5$$

$$= 25 + 15 - 33.5$$

$$= 6.5$$

$$U_2 = 5\times5 + \frac{5\times6}{2} - 21.5$$

$$= 25 + 15 - 21.5$$

$$= 18.5$$

$\therefore U_0 = 6.5$

<u>Critical value</u>: We have $\alpha = 0.05$ and $(n_1, n_2) = (5,5)$

$\therefore U_{(5,5),\ 0.05}$ (one-tailed test) $= 4$

<u>Decision</u>: Since $U_0 > U_{(5,5),\ 0.05}$. Hence, ~~there~~ the mean heart beating rate of non-vegetarian is not significantly high than that of vegetarian.

.1111

.1548

.13995 > 0.05

vegetarian

2

1

8.5

4

6

$R_2 = 21.5$

[LARGE SAMPLE SIZE]

For large sample size

$U_0 \sim N(\mu_{U_0}, \sigma^2_{\mu_0})$

Large sample size i.e. $n_1 > 10$ and $n_2 > 10$

For large sample size, the distribution of U_0 is approximately by normal distribution with mean $= \frac{n_1 n_2}{2}$

and variance $= \frac{n_1 n_2 (n_1+n_2+1)}{12}$

Test statistic: Under H_0

$$Z = \frac{U_0 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

In case of tied observation, the corrected value of s.d. is calculated as

$$\sigma_U = \sqrt{\frac{n_1 n_2}{n(n-1)} \left[\frac{n^3 - n}{12} - \frac{\Sigma (t_i^3 - t_i)}{12} \right]}$$

where, $n = n_1 + n_2$

t_i = number of times i^{th} rank is repeated

$X \sim N(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma^2}$$

$U_0 \sim N\left(\frac{n_1 n_2}{2}, \frac{n_1 n_2 (n_1+n_2+1)}{12}\right)$

$$Z = \frac{U_0 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

P 123

Q16

Solution

$n_1 = 11$

$n_2 = 9$

H_0: $Md_1 = Md_2$ i.e. yields of wheats per unit area is not different

H_1: $Md_1 \neq Md_2$ i.e. yields of wheats per unit area is different (Two-tailed test)

H_0: $Md_1 = Md_2$ i.e. there is no significant difference in the average yields between two varieties of wheat I and wheat II.

H_1: $Md_1 \neq Md_2$ i.e. there is significant difference in the average yields between two varieties of wheat I and wheat II (two-tailed test).

Calculation :

Wheat I	Wheat II	Ranks of wheat I (combined)	Ranks of
~~15.9~~	~~16.4~~	9	12.5
~~15.3~~	~~16.8~~	5	15
~~16.4~~	~~17.1~~	12.5	17
~~14.9~~	~~16.9~~	3	16
~~15.3~~	~~18.0~~	5	19
~~16.0~~	~~15.6~~	10.5	8
~~14.6~~	~~18.1~~	2	20
~~15.3~~	~~17.2~~	5	18
~~14.5~~	~~15.4~~	1	7
~~16.6~~		14	
~~16.0~~		10.5	
$t_1 = 3$		$R_1 = 77.5$	$R_2 =$
$t_2 = 2$			

Now,

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= 11\times 9 + \frac{11\times 12}{2} - 77.5$$

$$= 87.5$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$= 11\times 9 + \frac{9\times 10}{2} - 132.5$$

$$= 99 + 45 - 132.5$$

$$= 11.5$$

$\therefore U_0 =$ minimum of $\{U_1 \ \& \ U_2\}$

$$= 11.5$$

<u>Test-statistic</u>: Since, the problem is large sample size so

$$Z = \frac{U_0 - E(U_0)}{\sqrt{Var(U_0)}}$$

Now, $E(U_0) = \frac{n_1 n_2}{2} = \frac{11(9)}{2} = 49.5$

and $V(U_0) = \frac{n_1 n_2}{n(n-1)}\left[\frac{n^3-n}{12} - \frac{(3^3-3)+(2^3-2)+(2^3-2)}{12}\right]$

$$= 172.46$$

$\therefore Z = \frac{11.5 - 49.5}{\sqrt{172.46}} = -2.89 \Rightarrow |Z\text{-cal}| = 2.89$

Critical value: We have, $\alpha = 0.01$

$Z_{tab} = Z_{0.01}$ (two-tailed test) $= 2.575$

Decision: Since $|Z_{cal}| > Z_{tab}$, so we reject H_0.

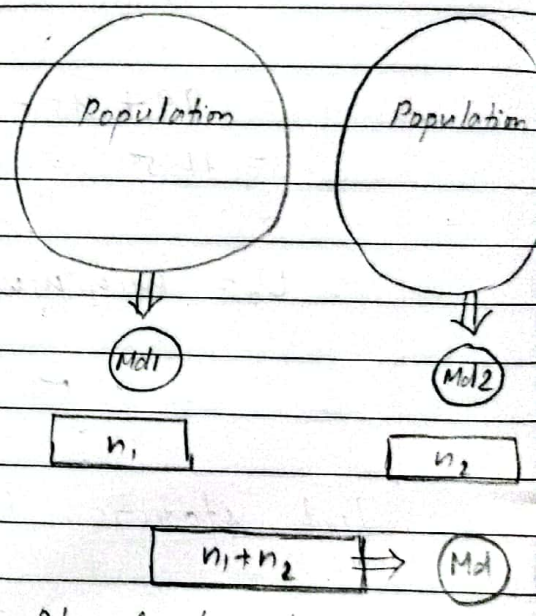
Assignment Q18

August 15, 2024

Median Test

Median Test:

Median test is non-parametric test used to test the difference in medians of two independent distributions. In other words, this test is applied to test whether two independent random samples have been drawn from two populations with same median or not. Also, it is used to test whether the two treatments applied in an experiment are equally effective or not.



No. of observations $\leq$ Md in the first sample

(a) Test statistic

Case I:

Small sample size i.e. $n_1 \leq 10$ and $n_2 \leq 10$.

Null Hypothesis (H_0): $Md_1 = Md_2$

i.e. there is no significant difference between two samples medians.

Alternative Hypothesis (H_1): $Md_1 \neq Md_2$ i.e. there is a

significant difference between two sample medians. (Two Tailed test)

or $H_1: Md_1 > Md_2$
or $H_1: Md_1 < Md_2$ } One Tailed test.

<u>Test statistic</u>: Under H_0, the test-statistic is obtained as follows:

(i) Combined ~~with~~ both the samples and arrange them in ascending order of magnitude such that $n = n_1 + n_2$.

(ii) Calculate the median of the combined sample and count the number of observations less or equal to median in the first sample which is called the test statistic and is denoted by 'a'.

Now,

$$P_0 = P(A = a)$$

$$= \frac{{}^{n_1}C_a \; {}^{n_2}C_{K-a}}{{}^{n_1+n_2}C_K}$$

where, $a = 0, 1, 2, \ldots,$ ~~min{(n, k)}~~ minimum of (n_1, K)

$$K = \frac{n_1 + n_2}{2}$$ ← round up when in decimal

<u>Critical value</u>:

We can obtain P-value (P_0) the probability associated with the ~~extreme~~ value as extreme as observed 'a' for n_1 and n_2.

$$\therefore P_0 = P(A \geq a)$$

<u>Decision</u>

For one-tailed test, P-value $= P_0 > \alpha$ then

we do not reject H_0 otherwise reject H_0. For two-tailed test, P-value $= 2P_0 > \alpha$, then we do not reject H_0 otherwise reject H_0.

Q7.

$n_1 = 10$ and $n_2 = 10$

H_0 : $Md_1 = Md_2$ i.e. the two training programs are equally effective.

H_1 : $Md_1 \neq Md_2$ i.e. the two training programs are not equally effective. (Two-tailed test)

<u>Test-statistic</u> : Under H_0, the test statistic is 'a'.

<u>Calculation</u>:

Arranging observations of both samples in ascending order of magnitude.

30, 32, 34, 39, 40, 42, 43, 44, 45, 46, 47, 47, 48, 49, 50, 51, 55, 55, 59, 71

(10) stm	(1) leaf
3	2 9 0 4
4	6 5 2 8 9 4
5	1 7 7 8
6	
7	1

$Md =$ Value of $\left(\frac{n+1}{2}\right)^{th}$ term ; $n = n_1 + n_2$

$= \frac{20+1}{2}$

$= 10.5^{th}$ term

$\therefore Md = \frac{46+47}{2} = 46.5$

Now, a = no. of observations less or equal to Md (46.5) in the first sample.

$= 7$

Calculation of P-value

$$P_0 = P(A \geq a) \; ; \; K = \frac{n_1 + n_2}{2} = \frac{10+10}{2} = 10$$

$$= \sum_{a=7}^{10} \frac{{}^{n_1}C_a \; {}^{n_2}C_{K-a}}{{}^{n_1+n_2}C_K}$$

$$= \frac{1}{{}^{20}C_{10}} \left[{}^{10}C_7 \; {}^{10}C_{10-7} + {}^{10}C_8 \; {}^{10}C_{10-8} + {}^{10}C_9 \; {}^{10}C_{10-9} + {}^{10}C_{10} \; {}^{10}C_{10-10} \right]$$

$$= \frac{16526}{184756}$$

$$= 0.0894$$

<u>Decision</u> : P-value $= 2P_0 = 2 \times 0.089$
$= 0.178$ [$\because$ The problem is two-tailed]

and $\alpha = 0.05$

Since, P-value $> \alpha$, so we do not reject H_0.
Hence, both the training programs are equally effective.

~~177~~

Assignment 10

Large Sample size ($n_i > 10$ and $n_2 > 10$)

	No. of obs. $\leq$ Md	No. of obs. $>$ Md	Total
Ist Sample	a	b	a+b
IInd Sample	c	d	c+d
Total	a+c	b+d	a+b+c+d = n

In this case, the median test is equivalent to χ^2-test.

Therefore, the test-statistic is

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2(1)$$

Critical value: The critical value of the test statistic is obtained from χ^2-table at α level of significance and 1 dof.

$$\chi^2_{tab} = \chi^2_{\alpha,1}$$

Decision: If $\chi^2_{cal} \geq \chi^2_{tab}$, then we reject H_0

otherwise we do not reject H_0.

Q9.

$n_1 = 12, \; n_2 = 12$

$n = n_1 + n_2 = 12 + 12 = 24$

H_0: $Md_1 = Md_2$ i.e. the marks distb^n of two teachers do not differ significantly.

H_1: $Md_1 \neq Md_2$ i.e. the marks distb^n of two teachers differ significantly (Two-tailed test).

Test statistic: Under H_0,

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Calculation:

Arranging observations of both samples in ascending order of magnitude:

31, 36, 43, 45, 49, 60, 62, 66,

73, 73, 74, 76, 77, 77, 77, 78, 78, 80, 82,

85, 88, 89, 91, 92

S	L
3	61
4	358
5	
6	602
7	3378478 67
8	92085

Assignment 8.10

Mod = value of $\left(\frac{n+1}{2}\right)^{th}$ term ; $n = n_1 + n_2$
$= 12 + 12 = 24$

$= \left(\frac{24+1}{2}\right)^{th}$ term

$= 12.5^{th}$ term

$\therefore$ Mod $= \frac{76+77}{2}$

$= 76.5$

2x2 Contingency table

	$x_i \le$ Mod	$x_i >$ Mod	Total
Sample I	a = 7	b = 5	12
Sample II	c = 5	d = 7	12
Total	(a+c) = 12	(b+d) = 12	24

$$\therefore \chi^2 = \frac{24(7\times7 - 5\times5)^2}{12\times12\times12\times12}$$

$= 0.666$

Critical value:

We have, $\alpha = 0.05$ and degree of freedom (df) = 1

$\therefore \chi^2_{tab} = \chi^2_{1, 0.05} = 3.841$

Decision: Since $\chi^2_{cal} < \chi^2_{tab}$, so we do not reject H_0. Hence, the marks distb^n of two teachers do not differ significantly.

August 17, 2020

Kolmogorov Smirnov Test (K-S Test)
[One sample, two samples]

χ^2 - test for the goodness-of-fit.

Kolmogorov-Smirnov (K-S) test is a test of goodness-of-fit. It is alternative to Chi-square test for goodness-of-fit when sample size is small.

Expected = $x \cdot P(x)$

$f(x) =$

$P(x) =$

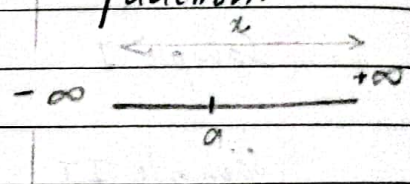
$F(x)$ = Cumulative probability distribution function.

One-Sample

Procedure:

H_0: $F(x) = F_0(x)$ i.e. there is no significant difference between observed and expected frequency. or frequencies are uniformly distributed.

H_1: $F(x) \neq F_0(x)$ i.e. there is significant difference between observed and expected frequency or frequencies are not uniformly distributed (Two Tailed test)



$F(a) = P(X \leq a)$

$f(x) = \frac{dF(x)}{dx}$

or H_1: $F(x) > F_0(x)$
or H_1: $F(x) < F_0(x)$ } (One Tailed Test).

Test-statistic : Under H_0,

$$D_0 = \text{Maximum } |F_e(x) - F_0(x)|$$

where, $F_e(x) = \frac{cfe}{n}$; cfe = expected cumulative frequency

$n = \Sigma f$ = Total frequency / Sample size.

fe = expected frequency $= n p_i$; p_i = probability of each category

$= \frac{\Sigma f}{\text{no. of categories}}$

$n \le 40$ Small
$n > 40$ large

$F_0(x) = \frac{cf_0}{n}$; cf_0 = Observed cumulative frequency

<u>Critical value</u>: The critical value of the test statistic is obtained from K-S table at α level of significance and size of sample 'n'.

$$\therefore D_{tab} = D_{n,\alpha}$$

<u>Decision</u>: If $D_0 \geq D_{n,\alpha}$ then we reject H_0 otherwise do not reject H_0.

Q17

H_0: The computers of five hard disk are uniformly infected.

H_1: The computers of five hard disk are not uniformly infected. (Two Tailed test).

<u>Test - statistic</u>: Under H_0,

$$D_0 = \text{Maximum} \left| F_e(x) - F_0(x) \right|$$

<u>Calculation</u>

Capacity of hard disk	No. of virus infected (O)	cf_0	$F_0(x)$	Expected (E)
500	11	11	11/50	
320	15	26	26/50	
1000	20	46	46/50	
2000	3	49	49/50	
4000	1	50	50/50	

$\longrightarrow$

Calculation:

Capacity of hard disk	No. of viruses infected (O)	cfo	$F_o(x)$	(E) = np	cfe	$F_e(x)$
500	11	11	11/50	10	10	10/50
320	15	26	26/50	10	20	20/50
1000	20	46	46/50	10	30	30/50
2000	3	49	49/50	10	40	40/50
4000	1	50	50/50	10	50	50/50

$\therefore D_0 = 16/50 = 0.32$

Critical value: We have, $\alpha = 0.05$ and $n = 50$

$\therefore D_{n,\alpha}$ (Two Tailed Test) $= D_{50, 0.05}$

$= \frac{1.36}{\sqrt{50}}$

$= 0.192$

Decision: Since $D_0 > D_{n,\alpha}$, we reject H_0. Hence, the computers of five hard disk are not uniformly infected.

Assignment

Q. A random sample of 20 volume based internet connected have following speed of internet connection in mps:

2.7, 2.9, 3.0, 3.1, 2.8, 3.0, 2.9, 3.0, 2.6, 3.1, 3.2, 3.1, 3.0, 2.9, 3.3, 3.0, 2.8, 2.9, 3.0, 2.9.

Apply Kolmogorov Smirnov test for testing that the internet speed are equally distributed, use $\alpha = 5\%$.

2.6 1
2.7 1
2.8 2
2.9 5

[a]
1/50
6/50
16/50
9/50
0

Two Samples Test:

[Small Sample Size, $n_1 = n_2 \leq 40$ or $n_1 \neq n_2 \leq 20$]

[Large Sample size, $n_1 = n_2 > 40$ or $n_1 \neq n_2 > 20$]

Null Hypothesis:

H_0: $F(x) = F(y)$ i.e. the two samples are drawn from same population.

H_1: $F(x) \neq F(y)$ i.e. the two samples are not drawn from same population. (Two Tailed test)

or H_1: $F(x) > F(y)$
or, H_1: $F(x) < F(y)$ } (One Tailed Test)

Small Sample Size

Test-statistic: Under H_0,

$$D_0 = \text{maximum} \left| F(x) - F(y) \right|; \quad F(x) = \frac{cf_x}{n_1}, \quad F(y) = \frac{cf_y}{n_2}$$

Critical value: The critical values of the test-statistic is obtained from K-S table at α level of significance for sample sizes n_1 and n_2.

$$D_{tab} = D_{(n_1, n_2), \alpha}$$

Q11. H_0: There is no significant difference between two types of motors.

H_1: There is significant difference between two types of motors. (Two Tailed Test)

<u>Test-statistic</u>: Under H_0,

$$D_0 = \text{Max} \, |F(x) - F(y)|$$

diff in years	f_x	f_y	cf_x	$F(x)$	cf_y	$F(y)$	$\lvert F(x) \rvert$
2	0	1	0	0/9	1	1/9	1/
3	3	2	3	3/9	3	3/9	0
4	2	2	5	5/9	5	5/9	0
5	2	1	7	7/9	6	6/9	1/
6	1	2	8	8/9	8	8/9	0
7	1	1	9	9/9	9	9/9	0
	$n_1 = 9$	$n_2 = 9$					

$$\therefore D_0 = \frac{1}{9}$$

$\therefore$ <u>Critical value</u>: We have, $\alpha = 0.05$, $n_1 = 9$, $n_2 = 9$

$$\therefore D_{(n_1, n_2), \alpha} = D_{(9,9), 0.05} \text{ (two-tailed test)}$$

$$= 5/9$$

<u>Decision</u>: Since $D_0 < D_{n_1, n_2, \alpha}$, we don't reject H_0. Thus, the two samples are drawn from same population. (Or there is no significant difference between two motors).

Aug 22, 2024

Large sample size

(i) $n_1, n_2 > 40$

$(n_1 = n_2)$

(ii) $n_1, n_2 > 20$

where, $n_1 \neq n_2$

Large Sample Size:

Test-Statistic:

$D_0 =$ maximum of $|F(x) - F(y)|$ [Two-tailed test]

$\chi^2 = 4D_0^2 \left(\frac{n_1 n_2}{n_1 + n_2}\right)$ [For one-tailed test]

Critical value:

For one tailed test, the critical value of the test-statistic is obtained by χ^2-table at α level of significance and 2 degree of freedom.

Decision

If $D_0 > D_{\alpha, (n_1, n_2)}$, then we reject H_0 Otherwise we do not reject H_0.

If $\chi^2_{cal} \geq \chi^2_{(\alpha, 2)}$, then we reject H_0 otherwise we do not reject H_0.

Q13. $n_1 = 26$, $n_2 = 25$ (large Sample size)

H_0: There is no significant difference between junior and senior programmers. $F(x) = F(y)$

H_1: There is a significant difference between junior and senior programmers. (Two Tailed Test) $F(x) \neq F(y)$

Test - Statistic: Under H_0,

$$D_0 = \text{Max} \, | F(x) - F(y) |$$

Calculation

No. of cigarettes	no. of junior prog. (f_x)	no. of senior prog. (f_y)	cf_x	$F(x)$	$F(y)$	$\lvert F(x) - F(y) \rvert$
0-2	7	5	7	7/26	5/25	0.069
2-4	6	4	13	13/26	9/25	0.14 ←
4-6	4	6	17	17/26	15/25	0.053
6-8	2	3	19	19/26	18/25	0.010
8-10	2	4	21	21/26	22/25	0.072
10-12	3	2	24	24/26	24/25	0.036
12-14	2	1	26	26/26	25/25	0
	$n_1 = 26$	$n_2 = 25$				

$\therefore D_0 = 0.14$

Critical value:

We have $\alpha = 0.05$, $n_1 = 26$, $n_2 = 25$ and the problem is two-tailed,

$$\therefore D_{0.05, (26, 25)} = 1.36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$= 1.36 \sqrt{\frac{26 + 25}{26 * 25}}$$

$$= \underline{0.38}$$

Decision

S

we do

Hen

junior

Q. Given

employ

company

Income

20 - 3

30 - 4

40 - 5

50 - 6

60 - 7

70 - 8

80 - 9

Decision :

Since $D_0 = 0.14 < D_{0.05, (26, 25)} = 0.30$ so, we do not reject H_0.

Hence, there is no significant difference between junior and senior programmers.

$F(y)$	$\lvert F(x) - F(y) \rvert$
5/25	0.069
9/25	0.14 ←
17/25	0.053
19/25	0.010
22/25	0.072
24/25	0.036
25/25	0

Q. Given below represents monthly income distribution of employees in a hardware company and software company.

Income ('000 Rs)	No. of employees in hardware company	No. of employees in software company
20 - 30	6	12
30 - 40	10	18
40 - 50	11	16
50 - 60	13	12
60 - 70	25	10
70 - 80	15	12
80 - 90	10	10

Scanned with CamScanner

Assignment

Do ~~is~~ the income distⁿ support that income of employees in Hardware company is more than income of employees in software company? Use K-S Test at $\alpha = 0.05$.

$$\chi^2_{(2, 0.05)} = 5.991$$

Alternative test, when population is non-normal

Wilcoxon Test

Paired t-test

Treatment → Objective

Effectiveness

	Students	1	2	3	4	5	6
Before treatment →	Score I	✓	✓	✓	✓	✓	✓
After treatment →	Score II	✓	✓	✓	✓	✓	✓

$\mu_1 > \mu_2$

Direction: One Tailed Test.

Large sample: $n > 25$ $\quad n_e = n - t$ → (difference of ties) effective

Individual	X	Y	d	Rank of |d|	Rank with (+)	Rank with (−)
1	10	8	2	2.5	2.5	—
2	15	17	−2	2.5	—	2.5
3	12	18	−6	6	—	6
4	13	14	−1	1	—	1
5	16	19	−3	4	—	4
6	20	16	+4	5	5	—
					$S(+) =$	$S(-) =$

Test Statistic:

$$T = \min\{S(+), S(-)\}$$

~~Wrct.~~

Wilcoxon Matched Pair Signed Rank Test:

It is used to test the effectiveness of a treatment applied in an experiment.

It is non-parametric test used to compare two populations for which observations are paired.

Procedure

H_0: $Md_1 = Md_2$ i.e. There is no significant difference in the scores before and after the treatment.

H_1: $Md_1 \neq Md_2$ i.e. There is significant difference in the scores before and after the treatment.

(Two Tailed Test)

H_1: $Md_1 > Md_2$
or H_1: $Md_1 < Md_2$ } (One Tailed Test)

where, Md_1 = Median score before treatment.
Md_2 = Median score after treatment.

Test-Statistic: Under H_0,

$$T = \text{Minimum of } \{S(+), S(-)\}$$

where,

$S(+)$ = Sum of the ranks of difference with '+' sign.

$S(-)$ = Sum of the ranks of difference with '−' sign.

NOTE: Ranks of $|d|$

where,

d = difference.

Critical value:

The critical value of the test-statistic is obtained from Wilcoxon Matched Paired Signed Rank Test table at α level of significance and effective sample size (n_e).

$$T_{tab} = T_{\alpha, n_e}$$

where, n_e = effective sample size
$= n - t$
t = no. of differences with zero.

August 24

(P140)

Q8.

Let, X = Score before course
Y = Score after course

H_0: $Md_1 = Md_2$ i.e. the course is not beneficial

H_1: $Md_1 < Md_2$ i.e. the course is beneficial (One Tailed Test)

Test-statistic: Under H_0,
T = minimum of $\{S(+), S(-)\}$

Calculation

X	Y	d = X − Y	Rank of |d|	Rank with sign '+'	'−'
57	60	−3	4	—	4
80	90	−10	8	—	8
64	62	+2	3	3	—
70	70	0		—	—
90	95	−5	6.5	—	6.5
59	58	+1	1.5	1.5	—
76	80	−4	5	—	5
98	99	−1	1.5	—	1.5
70	75	−5	6.5	—	6.5
83	94	−11	9	—	9
				$S(+) = 4.5$	$S(-) = 40.5$

24 2024

$\therefore T = 4.5$

(Wilcoxon) This is MW test

If calculated ≤ Tabulated, then we reject H_0

Critical value

We have, $\alpha = 0.05$ and $n_e = n - t$

$= 10 - 1$

$= 9$

$\therefore T_{n_e, \alpha}$ (One Tailed Test) $= T_{9, 0.05}$ (One Tailed Test)

$= 8$

Decision : Since $T = 4.5 < T_{9, 0.05} = 8$, so we reject H_0.

Hence, the course is beneficial.

Q9 (IMP) Let, X = First attempt
Y = Second attempt

H_0: $Md_1 = Md_2$ i.e. The difference between scores is not significant.
H_1: $Md_1 \neq Md_2$ i.e. The difference between scores is significant (Two tailed test)

Test statistic: Under H_0,
T = minimum of $\{S(+), S(-)\}$

Calculation

X	Y	d = X − Y	Rank of |d|	Rank with sign +	Rank with sign −
470	510	−40	6		6
530	550	−20	4		4
610	600	+10.	1	1	
440	490	−50	7		7
600	585	+15.	2	2	
590	620	−30	5		5
580	598	−18.	3		3
				$S(+) = 3$	$S(-) = 25$

$\therefore T = 3$

Critical value: We have $\alpha = 0.05$ & $n_e = 7$
$\therefore T_{n_e, \alpha}$ (Two tailed Test) $= T_{7, 0.05}$ (Two tailed Test)

$= 2$

Decision: Since $T = 3 > T_{7, 0.05} = 2$, we accept H_0.
Thus, there is no difference in scores between attempts.

(1 Sure Q)

Assign ranks based on overall [illegible]

Kruskal Wallis H-Test (K-W One Way ANOVA test)

It is used to test the difference between the locations of three or more samples.

K = no. of samples ≥ 3

Procedure

H_0 : $Md_1 = Md_2 = \dots = Md_K$ i.e.
there is no significant difference between k-samples or k-populations.

	T_1	T_2	T_3	T_4	T_5	
S_1	13	12	13	14	–	R_1
S_2	15	12	13	13	19	R_2
S_3	10	12	14		18	R_3
S_4	15	14	8	9	12	R_4

$n_1 = 4$
$n_2 = 5$
$n_3 = 3$
$n_4 = 5$
$(K=4)$

H_1 : $Md_1 \neq Md_2 \neq \dots \neq Md_K$ i.e.
there is a significant difference between k-samples or k-populations.

Test-statistic : Under H_0,

$$H = \frac{12}{n(n+1)} \left(\sum_{i=1}^{K} \frac{R_i^2}{n_i} \right) - 3(n+1)$$

where, $n = n_1 + n_2 + \dots + n_K$

n_1 = size of first sample.

n_2 = size of second sample.

. . .

. . .

. . .

n_K = size of kth sample.

R_1 = Sum of the ranks given to first sample

.

.

.

R_K = Sum of the ranks given to k^{th} sample.

Rank ⇒ Combined rank (i.e. using all the observations).

If there is a tie in the observations, then the corrected value of the test-statistic is obtained as follows:

$$H_{corrected} = \frac{H}{C.F.}$$

$$C.F. = \text{Correction Factor} = 1 - \frac{\Sigma(t_i^3 - t_i)}{n^3 - n}$$

t_i = no. of times i^{th} rank is repeated.

$$\therefore H_{corr} = \frac{\frac{12}{n(n+1)}\left(\sum_{i=1}^{k} \frac{R_i^2}{n_i}\right) - 3(n+1)}{1 - \frac{\Sigma(t_i^3 - t_i)}{n^3 - n}}$$

Critical value:

Small sample size:

$n_i \leq 5$ and $k = 3$ and the value of test-statistic, we obtain the p-value.

Decision: If P-value $> \alpha$, then we do not reject H_0. Otherwise, reject H_0.

Large sample size: ($n_i > 5$ and $k > 3$)

The critical value of the test-statistic is obtained from χ^2-table at α level of significance and $(k-1)$ degree of freedom.

$$\therefore \chi^2_{tab} = \chi^2_{\alpha, (k-1)}$$

Decision:

If $H < \chi^2_{tab}$ then we do not reject H_0 otherwise

reject H_0.

August 27, 2024

Q14.

Kruskal Wallis H-Test

$K = 3, \quad n_i = 4, \quad \forall i = 1, 2, 3$

H_0: $Md_1 = Md_2 = Md_3$ i.e. All three fertilizers are equally effective.

H_1: $Md_1 \neq Md_2 \neq Md_3$ i.e. All three fertilizers are not equally effective. (or at least one of the fertilizer is different).

Small sample
$n_i \leq 5$ & $K = 3$
p-value

Large sample
$n_i > 5$ and $K > 3$
$\downarrow$
$\chi^2_{\alpha, (K-1)}$

Calculation

N	P	K	R_1	R_2	R_3
122 (12)	81 (8)	80 (6)	10	8	6
80 (6)	80 (6)	82 (9)	6	6	9
138 (1)	79 (4)	65 (2.5)	12	4	2.5
121 (11)	65 (2.5)	58 (1)	11	2.5	1
			$\Sigma R_1 = 39$	$\Sigma R_2 = 20.5$	$\Sigma R_3 = 18.5$
			$\Sigma R_1^2/n = \frac{39^2}{4}$		
			$= 380.25$	105.06	85.56

Combined Rank

$n = 12$

$t_1 = 2, \quad t_2 = 3$

Test-statistic:

$$H_{corr} = \frac{H}{CF}$$

$$H = \frac{12}{n(n+1)} \left(\sum_{i=1}^{n} \frac{R_i^2}{n} \right) - 3(n+1)$$

$$= \frac{12}{12 \times 13} (310.25 + 105.06 + 85.56) - 3 \times 13$$

$$= 4.91$$

$$C.F. = 1 - \left[\frac{(8-2)+(27-3)}{12^3 - 12}\right]$$

$$= 1 - \left(\frac{6+24}{12^3-12}\right)$$

$$= 1 - \left(\frac{30}{12^3-12}\right)$$

$$= 0.982$$

$$\therefore H_{corrected} = \frac{4.91}{0.98} = \underline{5.01}$$

<u>Critical value</u>:

For $n_1 = 4$, $n_2 = 4$, $n_3 = 4$ and $H = 5.01$,

$P = 0.097$

<u>Decision</u>: We have, $\alpha = 0.05$ and P-value $= 0.097$.

Since, P-value $> \alpha$ so we do not reject H_0.

Hence, all three fertilizers are equally effective.

If H_0 is true $\Rightarrow$ insignificant
If H_1 is true $\Rightarrow$ significant

Q13.

$H_0: Md_1 = Md_2 = Md_3$

$H_1: Md_1 \neq Md_2 \neq Md_3$

Calculation.

S (10)	L (1)
6	179
7	49226
8	8752409
9	417

Method I	94 (17)	88 (12)	91 (14)	74 (6)	87 (10)	97 (18)		$\Sigma R_1 = 84$
Method II	85 (11)	82 (9)	79 (7)	84 (8)	61 (1)	72 (5.5)	80 (3)	$\Sigma R_2 = 55.5$
Method III	89 (13)	67 (2)	72 (5.5)	76 (4)	69 (3)			$\Sigma R_3 = 31.5$

$k = 2$

$n = 18$

Test statistic : Under H_0

$$H_{corrected} = \frac{H}{CF}$$

$$H = \frac{12}{n(n+1)} \left(\sum_{i=1}^{3} \frac{R_i^2}{n_i} \right) - 3(n+1)$$

$$= \frac{12}{18 \times 19} \left(\frac{84^2}{6} + \frac{55.5^2}{7} + \frac{31.5^2}{5} \right) - 3 \times 19$$

$$= \cancel{11.971} \quad 6.666$$

$$C.F. = 1 - \frac{(3 - 2)}{18^3 - 18}$$

$$= 1 - \left(\frac{6}{18^3 - 18} \right)$$

$$= 0.9989$$

$$\therefore H_{corr} = \cancel{\frac{11.971}{0.9989}} \cancel{=} \cancel{11.984} \quad \frac{6.667}{0.998} = \underline{6.68}$$

<u>Critical value</u> : We have $K = 3$ and $\alpha = 0.05$

$$\therefore \chi^2_{(K-1), \alpha} = \chi^2_{2, 0.05} = 5.991$$

<u>Decision</u> : Since, $H_{corrected} = 6.68 > \chi^2_{2, 0.05} = 5.991$, we reject H_0. Hence, all three methods are not equally effective.

Assignment

(15), (16), (17)

Fr. Friedman Two-Way ANOVA Test

Friedman F-test (at least 3 samples, $K \geq 3$)

K = no. of samples

n = size of each sample

		BLOCKS				
		b_1	b_2	b_3	b_4	Sum
SAMPLE	K_1					
	K_2					
	K_3					

We have to find rank of each block separately (Ascending order)

Then, sum for each sample.

Procedure: $H_0: Md_1 = Md_2 = \dots = Md_K$

i.e. there is no significant difference between K-sample medians or K-samples are identical

$H_1: Md_1 \neq Md_2 \neq \dots \neq Md_K$ i.e. there is significant difference between K-sample medians or K-samples are not identical. (or at least one of the sample is different).

Test-statistic: Under H_0,

$$F_{r\,corr.} = \frac{12}{nK(K+1)} \left(\sum_{i=1}^{K} R_i^2 \right) - 3n(K+1)$$

If tied case occurs then the corrected value of the test-statistic is

$$F_{corr} = \frac{F_r}{CF} \; ; \quad C.F. = 1 - \frac{\sum (t_i^3 - t_i)}{n(K^3 - K)}$$

t_i = no. of times i^{th} rank is repeated.

Critical value:

[Small] For $[2 \leq n \leq 9$ and $K = 3]$ and $[2 \leq n \leq 4$ and $K = 4]$

the P-value (P) is obtained from Friedman probability

table.

[Large] For $n \geq 5$ and $K > 3$, the critical value is $\chi^2_{\alpha,(K-1)}$

Sum

August 29, 2024

[Q 18, 19, 20]

Q18. $K = 3$, $n = 4$

H_0: $Md_1 = Md_2 = Md_3$ i.e. there is no significant difference in the performances of schools with respect to pass percentage

H_1: $Md_1 \neq Md_2 \neq Md_3$ i.e. there is significant difference in the performances of schools with respect to pass percentage.

Calculation

	Grade				R_i	R_i^2
	I	II	III	IV		
Alpha	89 (3)	98 (3)	70 (3)	80 (3)	$R_1 = 12$	$R_1^2 = 144$
Sigma	45 (2)	76 (2)	40 (2)	55 (1)	$R_2 = 7$	$R_2^2 = 49$
Gamma	20 (1)	58 (1)	35 (1)	67 (2)	$R_3 = 5$	$R_3^2 = 25$
						$\sum R_i^2 = 218$

Test-Statistic: Under H_0,

$$F_r = \frac{12}{nK(K+1)} \left(\sum_{i=1}^{K} R_i^2 \right) - 3n(K+1)$$

$$= \frac{12}{4 \times 3 (3+1)} * \left(218 \right) - 3 \times 4 \times 4$$

$$= \frac{12}{12 \times 4} \times 218 - 48$$

$$= 6.5$$

Q20 Assignment

P-value : For $n = 40$, $K = 3$ and $F_r = 6.5$,
P-value = 0.042

Decision : We have $\alpha = 0.05$ and P-value = 0.042.
Since, P-value $< \alpha$ so we reject H_0.
Hence, there is significant difference.

Q19

$k = 4$, $n = 4$

H_0: $Md_1 = Md_2 = Md_3 = Md_4$

H_1: $Md_1 \neq Md_2 \neq Md_3 \neq Md_4$

Calculation

	Winter	Spring	Summer	Fall
A	92 (2)	112 (4)	94 (3)	77 (1)
B	9 (1)	11 (3)	10 (2)	15 (4)
C	58 (2)	71 (4)	57 (1)	62 (3)
D	19 (2.5)	26 (4)	19 (2.5)	18 (1)
R_i	$R_1 = 7.5$	$R_2 = 15$	$R_3 = 8.5$	$R_4 = 9$
R_i^2	$R_1^2 = \cancel{29.25}$ 56.25	$R_2^2 = 225$	$R_3^2 = 72.25$	$R_4^2 = 81$
$t_1 = 2$				$\Sigma R_i^2 = 434.5$

Test statistic : Under H_0,

$$F_r = \frac{12}{4 \times 4 \times (4+1)} * 434.5 - 3 \times 4 \times 5$$

$$= 5.175$$

$$F_{r\,corr} = \frac{F_r}{CF}$$

$$C.F. = 1 - \frac{2^3 - 2}{4(4^3 - 4)} = 0.975$$

$$\therefore F_{corr} = \frac{F_r}{CF}$$

$$= \frac{5.175}{0.975}$$

$$= 5.307.$$

<u>Critical value</u>: for $n = 4$, $K = 4$ and $F_{corr} = 5.307$,
P-value $= 0.158$

<u>Decision</u>: Since P-value $(0.158) > \alpha(0.05)$, we accept H_0. Thus, there is no significant difference.

Q. Three different advertising media TV, Radio and Newspaper are being compared to study their effectiveness in promoting sales of Wai-Wai noodles. Each advertising media is exposed for specified period of time and sales ('000 packets) from 10 stores located at different areas are recorded.

	Stores									
Advertising Media	A	B	C	D	E	F	G	H	I	J
TV	20	21	15	12	14	17	21	16	20	18
Radio	7	9	11	12	10	10	14	12	8	7
Newspaper	8	6	11	12	9	6	8	10	8	6

Are three advertising media equally effective, use Friedman two way ANOVA test.

<u>Solution</u>

H_0: $Md_1 = Md_2 = Md_3$ i.e. all three advertising media are equally effective

H_1: $Md_1 \neq Md_2 \neq Md_3$ i.e. all three media are not equally effective.

Calculation

TV	20	21	15	12	14	17	21	16	20	18	$R_1 = 15$	$R_1^2 = 841$
Radio	7	9	11	12	10	10	14	12	8	7	$R_2 = 9$	$R_2^2 = 324$
Newspaper	8	6	11	12	9	6	8	10	8	6	$R_3 = 13$	$R_3^2 = 169$

$\Sigma R_i^2 = 1334$

$t_1 = 2$

$t_2 = 3$

$t_3 = 2$

$$F_r = \frac{12}{10 \times 3 \times 4} * 1334 - 3 \times 10 \times 4$$

$= 13.4$

$$C.F = 1 - \frac{(2^3 - 2) + (3^3 - 3) + (2^3 - 2)}{10(3^3 - 3)}$$

$= 0.85$

September 10, 2024

* Hypothesis Testing

α

β

$(1-\beta)$

Simple: Hypo complete info

Composite: Incomplete

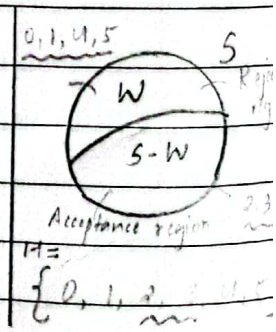
Simple Hypothesis

A hypothesis which completely specifies the distribution of random variables is known as simple hypothesis. For example, $X \sim N(\mu, \sigma^2)$ [i.e. X follows normal distribution with parameters μ and σ^2] and the hypothesis to be tested as $H_0: \mu = \mu_0, \sigma^2 = \sigma_0^2$, then under this hypothesis the form of the distribution is completely specified.

Composite Hypothesis

A hypothesis which ~~dis~~ does not completely ~~specified~~ specify the form of the distribution is said to be composite hypothesis. For example, $X \sim N(\mu, \sigma^2)$ and the hypothesis to be tested as $H_0: \mu = \mu_0$ then under this hypothesis σ^2 is not specified.

Critical Region



Let us define a Sample Space 'S' and we divide the whole sample space into two mutually exclusive (disjoint) subsets say, W and $S-W$ such that,

$$W \cup (S-W) = S$$

$$\text{and } W \cap (S-W) = \phi \quad (\text{Null set})$$

If the sample points fall in the region W, then we can reject the null hypothesis (H_0) and the region is called region of rejection or it is also called critical region and if the sample points fall in the region $(S-W$

then we accept the null hypothesis and the region is called region of acceptance.

Errors in Hypothesis Testing

Type I error:

Consumer's Risk

The error committed in rejecting H_0 when it is true is called Type I error. The probability of committing Type I error is called size of the test or size of the critical region and denoted by α and is given by

$\alpha = P(x \in W / H_0)$

$$\alpha = \text{Prob (type I error)}$$
$$= \text{Prob (reject } H_0 \text{ when } H_0 \text{ is true)}$$
$$= \text{Prob}(x \in W / H_0)$$
$$= \int_{x \in W} f(x/H_0)\, dx \quad [\text{Continuous}]$$
$$= \sum_{x \in W} P(x/H_0) \quad [\text{Discrete}]$$

Type II error: The error committed in accepting H_0 when H_0 is false called type second error. The probability of committing type II nd error is called size of the type II error and it is denoted by β and is given by:

Producer's Risk

$$\beta = \text{Prob (Type II error)}$$
$$= \text{Prob (accept } H_0 \text{ when } H_0 \text{ is false)}$$
$$= \text{Prob (accept } H_0 \text{ when } H_1 \text{ is True)}$$
$$= P(x \in S-W / H_1)$$
$$= \int_{x \in S-W} f(x/H_1)\, dx$$

Scanned with CamScanner

$$= \sum_{x \in S-W} P(x/H_1)$$

Power of the Test $(1-\beta)$

The complementary part probability of committing type II error is known as power of the test or power of the function and it is denoted by $(1-\beta)$ i.e.

$P(A/B) + P(A'/B) = 1$

$$1-\beta = 1 - \text{Prob}(\text{accept } H_0 \text{ when } H_0 \text{ is false})$$
$$= 1 - \text{Prob}(\text{accept } H_0 \text{ when } H_1 \text{ is true})$$
$$= \text{Prob}(\text{reject } H_0 \text{ when } H_1 \text{ is true})$$
$$= \text{Prob}(\text{accept } H_1 \text{ when } H_1 \text{ is true})$$
$$= \text{Prob}(x \in W/H_1)$$
$$= \int_{x \in W} f(x/H_1)\, dx \qquad \text{(Continuous)}$$

$$\Rightarrow (1-\beta) = \sum_{x \in W} P(x/H_1) \qquad \text{(Discrete)}$$

$$\Rightarrow (1-\beta) = \text{Prob}(\text{Correct decision})$$

Power of the test is also defined as the probability of correct decision.

In any test, we would like to make the probability of correct decision as large as possible. In fact, we would like to make $1-\beta = 1$ i.e. the probability of committing type II error as minimum as possible.

Q. Toss a coin 5 times, if 2 or 3 times head turn up then the coin is accepted as an unbiased coin. Find α, β and $(1-\beta)$ when $H_0: P = 1/2$ against $H_1: P = 3/4$. Find α, β.

$\alpha = \text{Prob}(x \in W/H_0)$

$\beta = \text{Prob}(x \in S-W/H_1)$

Solution

Let X = number of heads.

Since, a coin is tossed five times

$S = \{0, 1, 2, 3, 4, 5\}$

Now, $S - W = \{2, 3\}$

and $W = \{0, 1, 4, 5\}$

Here, $X \sim B(n, p)$ so $P(X = x) = {}^nC_x \, p^x \, q^{n-x}$; $x = 0, 1, \ldots$

We have, $n = 5$ and $H_0 : p = 1/2$ ~~and~~ vs $H_1 : p = 3/4$

$$\therefore \alpha = \text{Prob}\left(x \in W / H_0\right)$$

$$= P\left(x \in 0, 1, 4, 5 / p = 1/2\right)$$

$$= P\left(X = 0 / p = 1/2\right) + P\left(X = 1 / p = 1/2\right) + P\left(X = 4 / p = 1/2\right) + P\left(X = 5 / p = 1/2\right)$$

$$= {}^5C_0 (1/2)^0 (1/2)^{5-0} + {}^5C_1 (1/2)^1 (1/2)^{5-1} + {}^5C_4 (1/2)^4 (1/2)^{5-1} + {}^5C_5 (1/2)^5 (1/2)^{5-5}$$

$$= 0.375$$

$$\beta = P\left(x \in S - W / H_1\right)$$

$$= P\left(x \in 2, 3 / H_1\right)$$

$$= P\left(X = 2 / p = 3/4\right) + P\left(X = 3 / p = 3/4\right)$$

$$= {}^5C_2 (3/4)^2 (1/4)^{5-2} + {}^5C_3 (3/4)^3 (1/4)^{5-3}$$

$$= 0.087 + 0.263 = 0.35$$

Scanned with CamScanner

Sep 12, 2024

$\alpha = \text{Prob}(\text{reject } H_0 / H_0)$
$= P(x \in W / H_0)$

$\beta = \text{Prob}(\text{accept } H_0 / H_1)$
$= P(x \in \bar{W} / H_1)$

Q. If $X \sim P(\lambda)$ and critical region is defined as $X > 2$. To test the null hypothesis $H_0 : \lambda = 1$ against $H_1 : \lambda = 3$, find α, β and power of the test.

Solution:

Since, $X \sim P(\lambda)$ so $P(X = x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2, \ldots$ $\lambda > 0$

$X \sim P(\lambda)$
$P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2, \ldots$ $\lambda > 0$

Given, $W = \{x : x > 2\}$
$\bar{W} = \{x : x \leq 2\}$

Now, $P(X/H_0) = \dfrac{e^{-1}1^x}{x!}$ and $P(X/H_1) = \dfrac{e^{-3}3^x}{x!}$

Probability of type I error,

$$\alpha = P(x \in W / H_0)$$
$$= P(x > 2 / H_0)$$
$$= 1 - P(x \leq 2 / H_0)$$
$$= 1 - \left[P(X=0/H_0) + P(X=1/H_0) + P(X=2/H_0)\right]$$
$$= 1 - \left[\frac{e^{-1}1^0}{0!} + \frac{e^{-1}1^1}{1!} + \frac{e^{-1}1^2}{2!}\right]$$
$$= 1 - 0.919$$
$$= \cancel{0.0806}\ 0.0803$$

Scanned with CamScanner

Probability of type II error,

$$\beta = P\left(x \in s-w \,/\, H_1\right)$$

$$= P\left(x \le 2 \,/\, H_1\right)$$

$$= P\left(X=0 \,/\, H_1\right) + P\left(X=1 \,/\, H_1\right) + P\left(X=2 \,/\, H_1\right)$$

$$= P\left(X=0 \,/\, \lambda=3\right) + P\left(X=1 \,/\, \lambda=3\right) + P\left(X=2 \,/\, \lambda=3\right)$$

$$= \frac{e^{-3}\,3^0}{0!} + \frac{e^{-3}\,3^1}{1!} + \frac{e^{-3}\,3^2}{2!}$$

$$= \underline{0.423}$$

Power of the test,

$$1-\beta = 1-0.423$$

$$= \underline{0.576}$$

Q. If $x \ge 1$ is the critical region for testing $H_0: \theta = 2$ against $H_1: \theta = 1$ on the basis of single observation from the population,

$$f(x, \theta) = \theta e^{-\theta x} \; ; \; 0 < x \le \infty$$

Obtain the value of type I, type II errors. Also, find the power of the test.

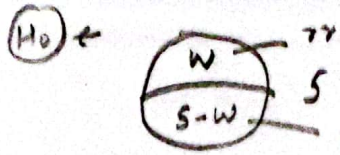
Solution

Given, $f(x, \theta) = \theta e^{-\theta x}$; $x > 0$

and $H_0: \theta = 2$ against $H_1: \theta = 1$

$$W = \{x : x \ge 1\} \text{ and } s-W = \{x : x < 1\}$$

Type I error



$\alpha = Prob(x \in W / H_0)$

$= \int_1^\infty f(x/H_0)\, dx$

$\Rightarrow \alpha = \int_1^\infty 2e^{-2x}\, dx$

$\Rightarrow \alpha = 2\left[\frac{e^{-2x}}{-2}\right]_1^\infty$

$\Rightarrow \alpha = -\left[e^{-2x}\right]_1^\infty$

$\Rightarrow \alpha = -\left(0 - \frac{1}{e^2}\right)$

$\Rightarrow \alpha = \frac{1}{e^2}$

$\therefore \alpha = \underline{0.135}$

Type II error

$\beta = Prob(x \in S-W / H_1) = P(x < 1 / H_1)$

Continuous so allowed

$\Rightarrow \beta = \int_0^1 f(x/H_1)\, dx$

$\int e^{ax} dx = \frac{e^{ax}}{a}$

$\Rightarrow \beta = \int_0^1 1e^{-x}\, dx$

$\Rightarrow \beta = \left[\frac{e^{-x}}{-1}\right]_0^1$

$\Rightarrow \beta = \left[e^{-x}\right]_1^0$

$\Rightarrow \beta = e^0 - \frac{1}{e}$ $\quad \therefore \beta = \underline{0.632}$

$\therefore$ Power of the test $= 1 - \beta = 1 - 0.632 = \underline{0.368}$

Scanned with CamScanner

Q. Given the freq" function,

$$f(x,\theta) = \begin{cases} 1/\theta \; ; & 0 \le x \le \theta \\ 0 \; ; & \text{otherwise} \end{cases}$$

and that you are testing null hypothesis $H_0 : \theta = 1$ against $H_1 : \theta = 2$.

What would be the size of the type I and type II errors, if you choose the interval

(i) $0.5 \le x$ (ii) $1 \le x \le 1.5$ as the critical regions?

Also obtain the power function of the test.

$W = \{x : x \ge 0.5\}$

Solution

$S-W = \{x : x < 0.5\} \; x \ge 0$

$$\alpha = \int_{x \in W} f(x/H_0)\, dx$$

$$\beta = \int_{x \in S-W} f(x/H_1)\, dx$$

$$\alpha = \int_{x \in W} f(x/H_0)\, dx$$

$$= \int_{0.5}^{\infty} f(x/H_0)\, dx$$

$$= \int_{0.5}^{1} 1/\theta \, dx + \cancel{\int_{1}^{\infty} 0 \, dx} \; 0$$

$$= \frac{1}{\theta} \left[x \right]_{0.5}^{1}$$

$$= \frac{1}{\theta} (1 - 0.5)$$

$$= \frac{1}{1} (0.5)$$

$$= 0.5$$

Scanned with CamScanner

Solution

Given $f(x, \theta) = \frac{1}{\theta}$; $0 \le x \le \theta$

$H_0 : \theta = 1$ vs $H_1 : \theta = 2$

(i) Given, $W = \{x : x \ge 0.5\}$

$S - W = \{x : x < 0.5\}$

Probability of type I error,

$$\alpha = \int_{x \in W} f\left(x / H_0\right) dx$$

$$= \int_{0.5}^{\theta} f\left(x / \theta = 1\right) dx$$

$$= \int_{0.5}^{1} 1/1 \, dx$$

Python

$$= \left[x\right]_{0.5}^{1}$$

$$= 1 - 0.5$$

$$= 0.5$$

$$\beta = \int_{x \in S-W} f\left(x / H_1\right) dx$$

$$= \int_{0}^{0.5} f\, x / \theta_1$$

$$= \int_0^{0.5} f\left(x/\theta=2\right) dx$$

$$= \int_0^{0.5} \frac{1}{2}\, dx$$

$$= \frac{1}{2}\left[x\right]_0^{0.5}$$

$$= \frac{1}{2} \times 0.5$$

$$= \underline{0.25}$$

Power of the test $= 1-\beta$

$= 1 - 0.25$

$= 0.75$

(ii) Given, $W = \{x : x \quad 1 \le x \le 1.5\}$

$S - W = \{x : x$

$\alpha = P\left(x \in W / H_0\right)$

$$= \int_{\boxed{1}}^{\boxed{1.5}} f\left(x/\theta=1\right) dx$$

x Doesn't lie in this interval

x is only defined for 0 to θ i.e. 0 to 1

$\Rightarrow \alpha = 0$ [Since probability does not exists) $H_0 : \theta = 1$ and ~~this~~ $f(x) = \frac{1}{\theta}$; $0 \le x \le \theta$]

$\beta = P\left(x \in S-W / H_1\right)$

$$= \int_{x \in S-W} f\left(x/H_1\right) dx = 1 - \int_{x \in W} f\left(x/H_1\right) dx$$

Scanned with CamScanner

NOTE: $\beta = P(x \in S-W \mid H_1)$

$\beta = 1 - P(x \in W \mid H_1)$

$$= 1 - \int_1^{1.5} f\left(x/_{\theta=2}\right) dx$$

$$= 1 - \int_1^{1.5} \frac{1}{2} \, dx$$

$$= 1 - \frac{1}{2}\left[x\right]_1^{1.5}$$

$$= 1 - \frac{(1.5-1)}{2}$$

$$= 1 - 0.25$$

$$= \underline{0.75}$$

$\therefore$ The power of the test,

$$1-\beta = 1 - 0.75 = \underline{\underline{0.25}}$$

Sep 14, 2024

Normal Distribution

If a continuous random variable X follows normal distribution with parameters μ and σ^2, then the probability density function of the distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} ; \quad \begin{cases} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{cases}$$

Standard Normal Distribution

If $X \sim N(\mu, \sigma^2)$ then $Z \sim N(0, 1)$ and then pdf of Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2 z^2} ; \quad -\infty < z < \infty$$

Scanned with CamScanner

where $z = \frac{x-\mu}{\sigma}$ = standard normal variate

$E(X) = \mu$

$V(X) = \sigma^2$

$E(Z) = 0$

$V(Z) = 1$

$V(aX) = a^2 V(X)$

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right]$$

$$= \frac{1}{\sigma}\left[E(X) - \mu\right]$$

$$= \frac{1}{\sigma}(\mu - \mu)$$

$$= 0$$

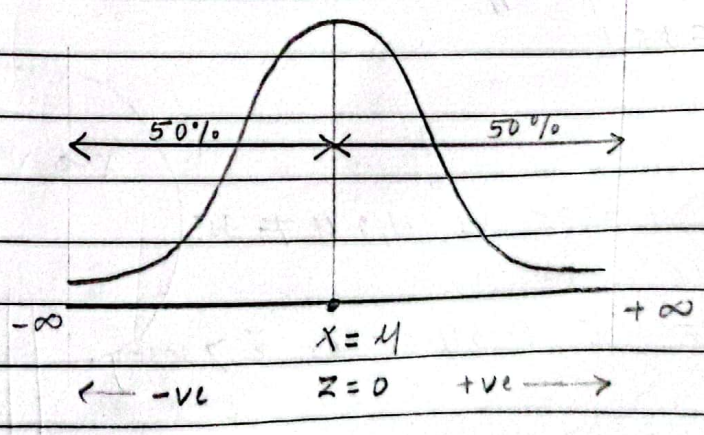
$$V(Z) = V\left[\frac{X-\mu}{\sigma}\right]$$

$$= \frac{1}{\sigma^2}\left[V(X) - V(\mu)\right]$$

$$= \frac{1}{\sigma^2}(\sigma^2 - 0)$$

$$= \sigma^2/\sigma^2 = 1$$

<u>Properties :</u>



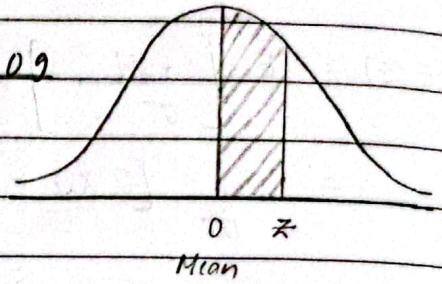
68.2%

95.4%

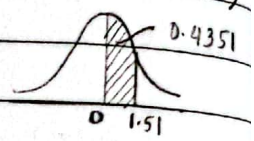
99.7%

Area under normal curve

z	0.00	0.01	...	0.09
0.0				
0.1				
.				
.		0.4345		
.				
3.6				



$P(0 < z < 1.51)$



Q. Suppose it is desired to test the hypothesis $H_0: \mu = 35$ against the alternative hypothesis $H_1: \mu \neq 35$ on the basis of random sample of size 16 from a normal pop^n $N(\mu, 1)$. The decision rule is; reject H_0 if the sample mean $\bar{x} < 34.65$ or $\bar{x} > 35.35$

(a) Find the prob of type I error

(b) Find the prob of type II error (i) when $\mu = 36$ and (ii) when $\mu = 36.1$

Solution

Given, $H_0: \mu = 35$ vs $H_1: \mu \neq 35$

$n = 16$

and $w = \{\bar{x} : \bar{x} < 34.65 \text{ or } \bar{x} > 35.35\}$

$\therefore S - w = \{\bar{x} : 34.65 \leq \bar{x} \leq 35.35\}$

$X \sim N(\mu, \sigma^2)$

n

$z = \frac{x - \mu}{\sigma}$

If $X \sim N(\mu, \sigma^2)$ then $\bar{x} \sim N(\mu, \sigma^2/n)$ $\quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Scanned with CamScanner

$\therefore \bar{x} \sim N\left(\mu, \frac{1}{n}\right)$ $\quad [\because X \sim N(\mu, 1)]$

(a) Probability of type I error

$$\alpha = Prob\left(x \in w / H_0\right)$$

$$= Prob\left(\bar{x} < 34.65 \text{ or } \bar{x} > 35.35 / \mu = 35\right)$$

$$\Rightarrow \alpha = P\left(\bar{x} < 34.65 / \mu = 35\right) + P\left(\bar{x} > 35.35 / \mu = 35\right)$$

$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$\Rightarrow \alpha = P\left(z < \frac{(34.65 - 35)}{1/\sqrt{16}}\right) + P\left(z > \frac{35.35 - 35}{1/\sqrt{16}}\right)$$

$$\Rightarrow \alpha = P(z < -1.4) + P(z > 1.4)$$

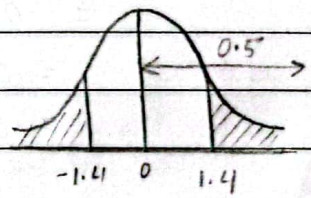
$$\Rightarrow \alpha = 2\,P(z > 1.4) \quad \text{(By symmetry)}$$

$$\Rightarrow \alpha = 2\left(0.5 - P(0 < z < 1.4)\right)$$

$$\Rightarrow \alpha = 2\,(0.5 - 0.4192)$$

$$\Rightarrow \alpha = 0.1616$$

$$\therefore \alpha = 0.1616$$

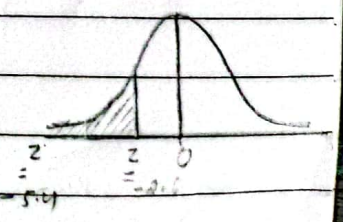


(b) Prob. of type II error when (i) $\mu = 36$ and (ii) $\mu = 36.1$

(i) $\mu = 36$

$$\beta = Prob\left(x \in s\text{-}w / H_1\right)$$

$$= P\left(\bar{x} : 34.65 \le \bar{x} \le 35.35 / \mu = 36\right)$$



$$= P\left(34.65 \le \bar{x} \le 35.35 / \mu = 36\right)$$

$$= P\left(\frac{34.65 - 36}{1/\sqrt{16}} \le z \le \frac{35.35 - 36}{1/\sqrt{16}}\right)$$

$-5.4 \qquad -2.6$

$$= P\left(-5.4 \le z \le -2.6\right)$$

$$= P\left(z < -2.6\right)$$

Scanned with CamScanner

$= 0.5 - P(-2.6 < Z < 0)$

$= 0.5 - 0.4953$

$= 0.0047$

(ii) $\mu = 36.1$

$\beta = Prob(x \in S\text{-}W / H_1)$

$= P(\bar{x} : 34.65 < \bar{x} < 35.35 \ (\mu = 36.1))$

$= P\left(\frac{34.65 - 36.1}{1/\sqrt{16}} < Z < \frac{35.35 - 36.1}{1/\sqrt{16}}\right)$

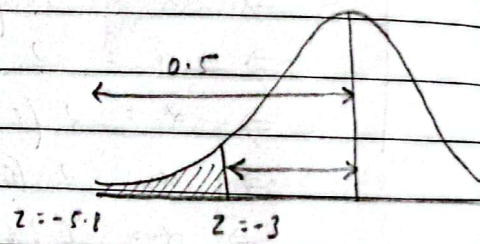
$= P(-5.8 < Z < -3)$

$= P(Z < -3)$

$= 0.5 - P(-3 < Z < 0)$

$= 0.5 - 0.4987$

$= 0.0013$



Scanned with CamScanner

[STAT] $\boxed{\text{iid.}}$ Sep 21, 2024

Q. Let, x_1, x_2, x_3, x_4, x_5 be a random sample of size 5 drawn from a $N(\mu, 4)$, μ unknown. If you choose $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 \geq 0$ as a critical region for testing $H_0: \mu = -1$ against $H_1: \mu = 1$, find the size of the critical region and power of the test.

$E(X_1 + X_2) = E(X_1) + E(X_2)$

$V(X_1 \pm X_2) = V(X_1) + V(X_2) \pm 2\,Cov(X_1, X_2)$

Solution

Given $x_i \sim N(\mu, 4)$ $\forall\, i = 1, 2, 3, 4, 5$

Let $y = x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5$

$$E(y) = E(x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5)$$

$$= E(x_1) + 2E(x_2) + 3E(x_3) + 4E(x_4) + 5E(x_5)$$

$$= \mu + 2\mu + 3\mu + 4\mu + 5\mu$$

$$= 15\mu$$

and $V(y) = V(x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5)$

$$= V(x_1) + 4V(x_2) + 9V(x_3) + 16V(x_4) + 25V(x_5)$$

$$= 4 + 4(4) + 9(4) + 16(4) + 25(4)$$

$$= 220$$

$\therefore y \sim N(15\mu, 220)$

Hence, $w = \{y : y \geq 0\}$ $\quad$ $s - w = \{y : y < 0\}$

$$\boxed{\text{Now, } Z = \frac{y - \mu}{\sigma}}$$

Prob. of type I error,

$$\alpha = Prob(x \in w / H_0)$$

$$= P(y \geq 0 / \mu = -1)$$

$$= P\left(Z \geq \frac{0 - 15(-1)}{\sqrt{220}}\right)$$

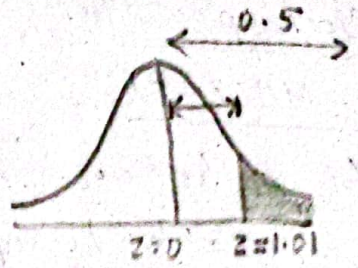
Scanned with CamScanner

$$\Rightarrow \alpha = P(Z \geq 1.011)$$

$$= 0.5 - P(0 \leq Z \leq 1.011)$$

$$= 0.5 - 0.3438$$

$$= \underline{0.1562}$$

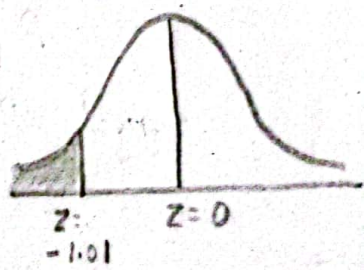


Probability of type II error,

$$\beta = Prob(x \in s\text{-}w / H_1)$$

$$= Prob(y < 0 / \mu = 1)$$

$$= P\left(Z < \frac{0-15}{\sqrt{220}}\right)$$



$$= P(Z < -1.011)$$

(By symmetry)

$$= P(Z > 1.011) = 0.5 - 0.3438$$

$$= \underline{0.1562}$$

β = minimum

$(1-\beta)$ = maximum

Best Critical Region

A critical region of low size α is said to be ~~the~~ a best critical region if it has minimum β or maximum $(1-\beta)$ among all critical regions.

A test which minimizes β or maximizes the power $(1-\beta)$ for a desired low level of significance α, is known as best test.

Scanned with CamScanner

Most powerful critical region (MPCR) and most powerful test (MP test)

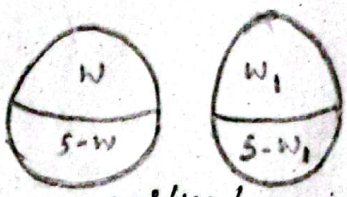
Let x be an arbitrary sample point in a sample space 'S'. Then, a critical region W is said to be a most powerful critical region (MPCR or BCR) of size α for testing a simple

Best Critical Region

null hypothesis $H_0: \theta = \theta_0$ against a simple alternative hypothesis $H_1: \theta = \theta_1$ if

$$P(x \in W / H_0) = \alpha \quad \text{——— (i)}$$

$$\text{and } P(x \in W / H_1) \geq P(x \in W_1 / H_1)$$



for every other critical region W_1 satisfying (i) and a test corresponding to this MPCR or BCR of size α is called most powerful (MP) test of size α.

(IMP)

Neyman Pearson Lemma (N-P Lemma)

Theorem

N-P lemma provides the general method of finding a most powerful test of simple hypothesis against the simple alternative hypothesis.

Statement:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size 'n' from popn with prob. function $f(x; \theta)$. Let K be any non-negative constant and W be a critical region for testing a simple hypothesis $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$, then the most powerful

Scanned with CamScanner

critical region (MPCR) W of size α (fixed) is given by

$$\frac{L(x/H_1)}{L(x/H_0)} \geq k \quad \forall\, x \in W$$

or, $$\frac{L(x/H_1)}{L(x/H_0)} < k \quad \forall\, x \in S-W$$

where, $L(x/H_0)$ and $L(x/H_1)$ are the likelihood functions of the sample observations under H_0 and H_1 respectively.

Proof : Since, α is the probability of ~~fix~~ type I error

$$\therefore \alpha = \text{Prob(Type I error)} = \int_{x \in W} L(x/H_0)\, dx \quad \text{—— (i)}$$

$$= P(x \in W / H_0)$$

and β is the probability of type II error,

$$\therefore \beta = \text{Prob(Type II error)}$$

$$\Rightarrow \beta = P(x \in S-W / H_1)$$

$$= \int_{x \in S-W} L(x/H_1)\, dx$$

$$\Rightarrow \beta = 1 - \int_{x \in W} L(x/H_1)\, dx \quad \text{—— (ii)}$$

Let W^* be any other critical region with prob. of type I error α_1 (which is less than or equal to α) and prob. of type II error β_1

$$\therefore \alpha_1 = \text{Prob (Type I error)}$$

$$= P(x \in W^* / H_0)$$

$$\Rightarrow \alpha = \int_{x \in W^*} L(x/H_0)\, dx \quad \text{—— (iii)}$$

Scanned with CamScanner

and β_1 = Prob (Type II error)

$$= \text{Prob}\left(x \in S - w^* / H_1 \right)$$

$$\Rightarrow \beta_1 = 1 - P\left(x \in w^* / H_1 \right) \quad \text{——— (iv)}$$

Since $\alpha \geq \alpha_1$

$$\Rightarrow \int_{x \in w} L\left(x/H_0\right) dx \geq \int_{x \in w^*} L\left(x/H_0\right) dx \quad \text{——— (v)}$$

Scanned with CamScanner

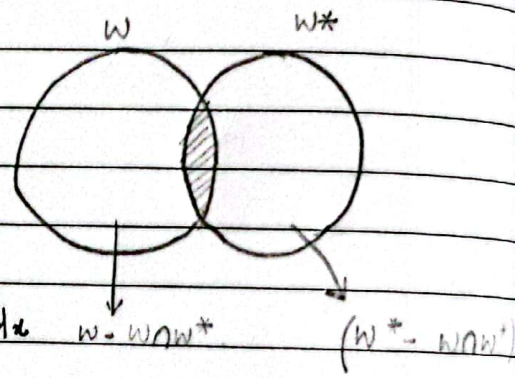
Oct 1, 2024

N-P lemma (Continue)

From eq^{n}s (ii) and (iv)

$$\beta_1 - \beta = 1 - \int_{x \in W^*} L(x/H_1)\,dx - 1 + \int_{x \in W} L(x/H_1)\,dx$$

$$\Rightarrow \beta_1 - \beta = \int_{x \in W} L(x/H_1)\,dx - \int_{x \in W^*} L(x/H_1)\,dx$$



$$\Rightarrow \beta_1 - \beta = \int_{x \in W - W \cap W^*} L(x/H_1)\,dx + \cancel{\int_{x \in W \cap W^*} L(x/H_1)\,dx} - \int_{x \in W^* - W \cap W^*} L(x/H_1)\,dx - \cancel{\int_{x \in W \cap W^*} L(x/H_1)\,dx}$$

$$\Rightarrow \beta_1 - \beta = \int_{x \in W - W \cap W^*} L(x/H_1)\,dx - \int_{x \in W^* - W \cap W^*} L(x/H_1)\,dx \quad \text{——— (v)}$$

From N-P lemma,

$$\frac{L(x/H_1)}{L(x/H_0)} > K \quad \forall\, x \in W$$

$1 - \beta_1 \leq 1 - \beta$

$$\Rightarrow L(x/H_1) \geq K\, L(x/H_0)$$

Now,

$$\int_{x \in W - W \cap W^*} L(x/H_1)\,dx \geq K \int_{x \in W - W \cap W^*} L(x/H_0)\,dx \quad \text{——— (vi)}$$

and

$$\int_{x \in W^* - W \cap W^*} L(x/H_1)\,dx \geq K \int_{x \in W^* - W \cap W^*} L(x/H_0)\,dx \quad \text{——— (vii)}$$

Scanned with CamScanner

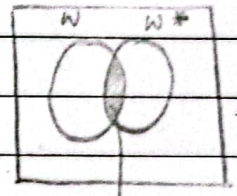
$$\Rightarrow \cancel{\beta_1 - \beta = \int_{x \in W - W \cap W^*} L(x/H_1)\,dx - \int_{x \in W^* - W \cap W^*} L(x/H_1)\,dx} \quad \text{— (vii)}$$

Using equations (vi) and (vii) in eqn (v), we get:

$$\beta_1 - \beta \geq K \int_{x \in W - W \cap W^*} L(x/H_0)\,dx - K \int_{x \in W^* - W \cap W^*} L(x/H_0)\,dx \quad \text{— (viii)}$$

From eqn (iii),

$$\int_{x \in W^*} L(x/H_0)\,dx \leq \int_{x \in W} L(x/H_0)\,dx$$



$$\Rightarrow \int_{x \in W - W \cap W^*} L(x/H_0)\,dx - \int_{x \in W^* - W \cap W^*} L(x/H_0)\,dx \geq 0 \quad \text{— (ix)}$$

$W \cap W^*$

$$\Rightarrow \cancel{\int_{x \in W - W \cap W^*} L(x/H_1)\,dx - \int_{x \in W^* - W \cap W^*} L(x/H_1)\,dx \geq 0}$$

From equation (ix) and (v), we get:

$$\beta_1 - \beta \geq 0$$

$$\Rightarrow \beta_1 \geq \beta$$

$$\Rightarrow 1 - \beta_1 \leq 1 - \beta$$

$$\boxed{\therefore \; 1 - \beta_1 \leq 1 - \beta}$$

$1 - \beta \geq 1 - \beta_1$. Hence, the critical region W of the fixed α is the most powerful critical region (MPCR) or BCR of the size α and the test is the most powerful test.

Scanned with CamScanner

Theorem

Statement: In testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_1$, the power of the MP critical region (MPCR or BCR) i.e. the power of the MP test is never less than the size of the BCR.

$$\therefore 1 - \beta \geq \alpha$$

Proof

Let w be a MP critical region or a BCR of size α (fixed) in testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, then by definition, we have,

$$\alpha = \int_{x \in w} L(x/H_0)\, dx \quad \text{and}$$

$$\beta = \int_{x \in S-w} L(x/H_1)\, dx$$

For N-P lemma, the BCR w is given by

$$\frac{L(x/H_1)}{L(x/H_0)} \geq k \quad \forall x \in w \Rightarrow L(x/H_1) \geq k\, L(x/H_0)$$

$$\Rightarrow \int_{x \in w} L(x/H_1)\, dx \geq k \int_{x \in w} L(x/H_0)\, dx$$

$$\Rightarrow (1-\beta) \geq k\alpha \quad \text{--- (i)}$$

Further,

$$\frac{L(x/H_1)}{L(x/H_0)} < k \quad \forall x \in S-w$$

$$\Rightarrow L(x/H_1) < k\, L(x/H_0)$$

$$\Rightarrow \int_{x \in S-w} L(x/H_1)\, dx < k \int_{x \in S-w} L(x/H_0)\, dx$$

Scanned with CamScanner

$\Rightarrow \beta \leq (1-\alpha) K$ — (ii)

$\Rightarrow (1-\alpha) K \geq \beta$

From (i) & (ii),

$(1-\beta)(1-\alpha) K \geq K \alpha \beta$

$\Rightarrow 1-\alpha-\beta+\alpha\beta \geq \alpha\beta$

$\Rightarrow 1-\alpha-\beta \geq 0$

$\Rightarrow \cancel{-\alpha-\beta \geq 0}$

$\boxed{1-\beta > \alpha}$

Proved

Hence, the power of the MP test is never less than the size of the MP critical region (BCR).

November 13, 2024

(IMP) Q. Let $x_1, x_2, \ldots, x_n$ be a random sample of size 'n' drawn from a normal population $N(\mu, \sigma^2)$ where σ^2 is known. Then for testing $H_0: \mu = \mu_0$ against $H_1: \mu = \mu_1$, find best critical region (BCR) (i) If $\mu_1 > \mu_0$ and (ii) If $\mu_1 < \mu_0$. Find power of the tests for the above two cases.

Solution If $X \sim N(\mu, \sigma^2)$ then the pdf the distⁿ is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; -\infty < x < \infty$$

Now, the likelihood function of $X = x_1, x_2, \ldots, x_n$

$\alpha = \text{Prob}(\text{Reject } H_0 / H_0 \text{ is true})$

$= P(x \in W / H_0)$

$\beta = \text{Prob}(\text{accept } H_0 / H_1 \text{ is true})$

$= P(x \in S-W / H_1)$

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$-\infty < x < \infty$

$-\infty < \mu < \infty$

Scanned with CamScanner

iid

$$L = \prod_{i=1}^{n} f(x_i)$$

$$= \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x_1-\mu}{\sigma}\right)^2} \right\} \cdots \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x_n-\mu}{\sigma}\right)^2} \right\}$$

$$\Rightarrow L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu)^2}$$

Now, $L\left(x/H_0\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu_0)^2}$ [$\because$ Under H_0, $\mu = \mu_0$]

and $L\left(x/H_1\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu_1)^2}$ [$\because$ Under H_1, $\mu = \mu_1$]

Now, from N-P lemma, the best critical region 'w' is given by:

$$\frac{L\left(x/H_1\right)}{L\left(x/H_0\right)} \geq k \quad \forall x \in w$$

$$\Rightarrow \frac{e^{-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu_1)^2}}{e^{-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu_0)^2}} \geq k$$

$$\Rightarrow e^{-\frac{1}{2\sigma^2}\left[\Sigma(x_i-\mu_1)^2 - \Sigma(x_i-\mu_0)^2\right]} \geq k$$

Taking log on both sides, we get:

$$\frac{-1}{2\sigma^2}\left[\Sigma(x_i-\mu_1)^2 - \Sigma(x_i-\mu_0)^2\right] \geq \log k$$

$$\Rightarrow \cancel{\frac{-1}{2\sigma^2}\Sigma(x_i-\mu_1+x_i-\mu_0)(x_i-\mu_1-x_i+\mu_0)}$$

$$\Rightarrow -\frac{1}{2\sigma^2}\left[\Sigma\left(x_i^2 - 2x_i\mu_1 + \mu_1^2\right) - \Sigma\left(x_i^2 - 2x_i\mu_0 + \mu_0^2\right)\right] \geq \log k$$

Scanned with CamScanner

$$\Rightarrow \frac{-1}{2\sigma^2}\left[\cancel{\Sigma x_i^2} - 2\mu_1(\Sigma x_i) + n\mu_1^2 - \cancel{\Sigma x_i^2} + 2\mu_0(\Sigma x_i) - n\mu_0^2\right] \geq \log k$$

$$\Rightarrow \frac{-1}{2\sigma^2}\left[-2n\mu_1\bar{x} + 2n\mu_0\bar{x} + n(\mu_1^2 - \mu_0^2)\right] \geq \log k$$

$$\Rightarrow \frac{-1}{2\sigma^2}\left[-2n\bar{x}(\mu_1 - \mu_0) + n(\mu_1^2 - \mu_0)^2\right] \geq \log k$$

$$\Rightarrow n\bar{x}(\mu_1 - \mu_0) - \frac{n}{2}(\mu_1^2 - \mu_0)^2 \geq \sigma^2 \log k$$

$$\Rightarrow n\bar{x}(\mu_1 - \mu_0) \geq \sigma^2 \log k + \frac{n}{2}(\mu_1^2 - \mu_0^2)$$

$$\Rightarrow \bar{x}(\mu_1 - \mu_0) \geq \frac{\sigma^2 \log k}{n} + \frac{(\mu_1^2 - \mu_0^2)}{2} \quad \text{——— (i)}$$

<u>Case I</u>

If $\mu_1 > \mu_0$, the BCR w is given by

$$\bar{x} \geq \frac{\sigma^2}{n}\frac{\log k}{(\mu_1 - \mu_0)} + \frac{(\mu_1 + \mu_0)}{2}$$

$$\Rightarrow \bar{x} \geq k_1$$

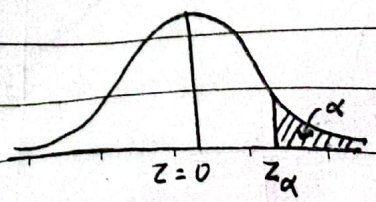
where, $k_1 = \frac{\sigma^2}{n}\frac{\log k}{(\mu_1 - \mu_0)} + \frac{(\mu_1 + \mu_0)}{2}$

$\therefore$ BCR is $w = \{x : \bar{x} \geq k_1\}$

Where the value of k_1 is obtained from fixed α

Now, $\alpha = \text{Prob}(x \in w / H_0)$

$\Rightarrow \alpha = \text{Prob}(\bar{x} \geq k_1 / H_0)$



Scanned with CamScanner

$\left[\bar{x} \pm Z_{\frac{\alpha}{2}} S.E.(\bar{x})\right]$

$$\Rightarrow \alpha = P\left(z \geq \frac{k_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\Rightarrow \frac{k_1 - \mu_0}{\sigma/\sqrt{n}} = z_\alpha$$

$$\Rightarrow k_1 = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Hence, the BCR, $w = \left\{ x : \bar{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\}$

Now, the power of the test

$$1-\beta = P(x \in w / H_1)$$

$$\Rightarrow 1-\beta = P\left(\bar{x} \geq k_1 / \mu = \mu_1\right)$$

$$\Rightarrow 1-\beta = P\left(z \geq \frac{k_1 - \mu_1}{\sigma/\sqrt{n}}\right)$$

$$\Rightarrow 1-\beta = P\left(z \geq \frac{\mu_0 + z_\alpha \sigma/\sqrt{n} - \mu_1}{\sigma/\sqrt{n}}\right)$$

$$\Rightarrow 1-\beta = P\left(z \geq z_\alpha - \frac{(\mu_1 - \mu_0)}{\sigma/\sqrt{n}}\right)$$

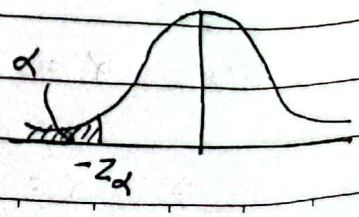
Case II :

$\mu_1 < \mu_0$ (left tailed), the BCR is given by

$$\bar{x} \leq \frac{\sigma^2}{n} \frac{\log k}{(\mu_1 - \mu_0)} + \cancel{\frac{(\mu_1 + \mu_0)}{2}} \quad \frac{(\mu_1 + \mu_0)}{2}$$

$$\Rightarrow \bar{x} \leq k_2$$

where $k_2 = \frac{\sigma^2}{n} \frac{\log k}{(\mu_1 - \mu_0)} + \frac{(\mu_1 + \mu_0)}{2}$



Scanned with CamScanner

$\therefore$ The BCR $w = \{ x : \bar{x} < k_2 \}$ and the value of k_2 is obtained from fixed α.

$$\alpha = Prob\left(x \in w / H_0 \right)$$

$$= Prob\left(\bar{x} \in k_2 / \mu = \mu_0 \right)$$

$$= Prob\left(Z \leq \frac{k_2 - \mu_0}{\sigma/\sqrt{n}} \right)$$

$$\Rightarrow \frac{k_2 - \mu_0}{\sigma/\sqrt{n}} = -Z_\alpha$$

$$\Rightarrow k_2 = \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$\therefore$ The BCR $(w) = \left\{ x : \bar{x} \leq \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} \right\}$

Now the power of the test:

$$1 - \beta = P\left(x \in w / H_1 \right)$$

$$\Rightarrow 1 - \beta = P\left(\bar{x} \leq k_2 / \mu = \mu_1 \right)$$

$$\Rightarrow 1 - \beta = P\left(Z \leq \frac{k_2 - \mu_1}{\sigma/\sqrt{n}} \right)$$

$$\Rightarrow 1 - \beta = P\left(Z \leq -Z_\alpha + \frac{(\mu_0 - \mu_1)}{\sigma/\sqrt{n}} \right)$$

$$\Rightarrow 1 - \beta = P\left(Z \leq \frac{(\mu_0 - \mu_1)}{\sigma/\sqrt{n}} - Z_\alpha \right)$$

Scanned with CamScanner

1%	5%	10%
$z_{0.01} = 2.33$	$z_{0.05} = 1.645$ (1T)	$z_{0.1} = 1.28$

3 marks

$1.96, 2.576$ (2T)

NOTE: When $\mu_1 < \mu_0$

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

and $\mu_1 < \mu_0$

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_1)^2}$$

Q. Let a random sample of size 40 is selected from a $N(\mu, 16)$. In testing $H_0: \mu = 50$ against $H_1: \mu = 52$ at 5% level of significance, find

(i) Best critical region of size α

(ii) Prob. of type II error

(iii) Power of the test

(iv) Also is this test unbiased?

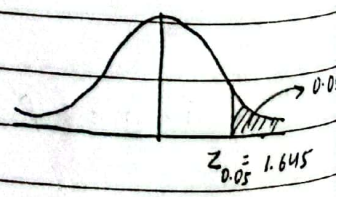
Solution

Given, $n = 40$, $\sigma^2 = 16$, $\alpha = 0.05$

For testing $H_0: \mu = 50$ against $H_1: \mu = 52$ (>50), the right tailed test at α level of significance, the BCR of size α is given by:

$$W = \{ x : \bar{x} \geq k_1 \}$$

where, $k_1 = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$



$$= 50 + 1.645 * \frac{4}{\sqrt{40}}$$

$$= 50 + 1.04$$

$$= \boxed{51.0411}$$

Scanned with CamScanner

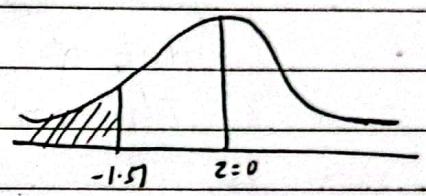
(ii) The prob. of type II error is given by,

$$\beta = \text{Prob}\left(x \in S-W / H_1\right)$$

$$= \text{Prob}\left(x < 57.0411 / \mu_1 = 52\right)$$

$$= \text{Prob}\left(z < \frac{57.0411 - 52}{4/\sqrt{40}}\right)$$

$$= \text{Prob}\left(z < -1.57\right)$$



$$= 0.5 - P(-1.5 < z < 0)$$

$$= 0.5 - 0.4357$$

$$= \boxed{0.0643}$$

(iii)

Power of the test $(1-\beta) = 1 - 0.0643$

$$= \boxed{0.9357}$$

(iv) Since, power of the test $(1-\beta) = 0.9357$ is very large as compared to $\alpha = 0.05$, so the test is unbiased.

November 14, 2024

Q. Suppose we want to test $H_0: \mu = 10$ against $H_1: \mu = 8$ using a random sample of size 'n' from a normal population $N(\mu, 4)$ and we reject the null hypothesis H_0 if $\bar{x} < c$. If $\alpha = 0.05$ and $\beta = 0.10$,

Scanned with CamScanner

find sample size 'n' and constant 'c'.

Solution

Given, $X_i \sim N(\mu, 4)$, $\forall\, i = 1, 2, \ldots, n$

$$\Rightarrow \bar{x} \sim N(\mu, 4/n)$$

$\alpha = 0.05$

$\boxed{Z_\alpha = -1.645}$

$\beta = 0.10$

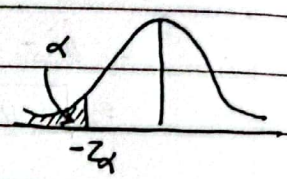
$\boxed{Z_\beta = 1.28}$

Here, we want to test $H_0: \mu = 10$ against $H_1: \mu = 8\ (< 10)$, left-tailed test at α level of significance.

The critical region, $W = \{x : \bar{x} < c\}$

$$\alpha = Prob\left(x \in W / H_0\right)$$

$$= P\left(\bar{x} < c / \mu = 10\right)$$



$$\Rightarrow \alpha = P\left(Z < \frac{c-10}{2/\sqrt{n}}\right)$$

$$= \frac{c-10}{2/\sqrt{n}}$$

$$\Rightarrow \frac{c-10}{2/\sqrt{n}} = -Z_\alpha$$

$$\Rightarrow c = 10 - \frac{2}{\sqrt{n}} Z_\alpha$$

$$\Rightarrow c = 10 - \frac{2}{\sqrt{n}}(1.645)$$

$$\therefore c = 10 - \frac{2}{\sqrt{n}}(1.645)$$

Now, the sample size,

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_0 - \mu_1)^2}$$

$$= \frac{(Z_{0.05} + Z_{0.1})\, 4}{(10-8)^2}$$

Scanned with CamScanner

$= 8.55$

$\therefore \boxed{n \approx 9}$ (Take ceiling value)

Now,

$$c = 10 - \frac{2}{\sqrt{9}}(1.645)$$

$$= 10 - \frac{2}{3} * (1.645)$$

$$= \boxed{8.90}$$

H/W Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a $N(\mu, 100)$ and $W = \{x : \bar{x} \geq c\}$ is a BCR of size α for testing $H_0 : \mu = 75$ against $H_1 : \mu = 78$ at α level. If the probability of type I and type II errors are fixed at $\alpha = 0.05$ and $\beta = 0.10$, find the sample size 'n' and the constant 'c'.

Ans
$n = 95$
$c = 76.69$

Q. Obtain BCR and power function of test for testing $H_0 : \theta = 1$ against $H_1 : \theta = \theta_1 (>1)$ on the basis of a sample size 1 from the population with pdf:

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

Solution

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

We have to test $H_0 : \theta = 1$ vs $H_1 : \theta = \theta_1 \ (>1)$.

Now, $f(x/H_0) = 1 x^{1-1} = 1$

$\theta_1 = 1 \quad ; \quad 0 < x < 1$

Scanned with CamScanner

Now, from NP-lemma, the best critical region is obtained as

$$\frac{f(x/H_1)}{f(x/H_0)} \geq k, \quad \forall x \in W$$

$$\Rightarrow \frac{\theta_1 x^{\theta_1 - 1}}{1} \geq k$$

$$\Rightarrow x^{\theta_1 - 1} \geq \frac{k}{\theta_1}$$

$$\Rightarrow \log\left(x^{\theta_1 - 1}\right) \geq \log\left(\frac{k}{\theta_1}\right)$$

$$\Rightarrow (\theta_1 - 1) \log x \geq \log\left(\frac{k}{\theta_1}\right)$$

$$\Rightarrow \log x \geq \frac{\log(k/\theta_1)}{(\theta_1 - 1)}$$

$$\Rightarrow x \geq \text{antilog}\left[\frac{\log(k/\theta_1)}{(\theta_1 - 1)}\right] = c \text{ (say)}$$

$$\therefore W = \{x : x \geq c\}$$

The value of 'c' can be determined from fixed α.

$$\therefore \alpha = \text{Prob}(x \in W / H_0)$$

$$\Rightarrow \alpha = P(x \geq c / H_0)$$

$$\Rightarrow \alpha = \int_c^1 f(x/H_0)\, dx$$

$$\Rightarrow \alpha = \int_c^1 1\, dx$$

$$\Rightarrow \alpha = [x]_c^1$$

$$\Rightarrow \alpha = 1 - c$$

$$\boxed{\therefore c = 1 - \alpha}$$

Scanned with CamScanner

$\therefore W = \{ x : x \geq (1-\alpha) \}$

Power of the test

$$\beta = P(x \in W / H_1)$$

$$= P(x \geq c / H_1)$$

$$= \int_c^1 f(x/H_1)\, dx$$

$$= \int_c^1 \theta_1 x^{\theta_1 - 1}\, dx$$

$$= \theta_1 \left[\frac{x^{\theta_1}}{\theta_1} \right]_c^1$$

$$= \left[x^{\theta_1} \right]_c^1$$

$$= 1^{\theta_1} - c^{\theta_1}$$

$$\therefore \boxed{\beta = (1 - c^{\theta_1})}$$

$$\therefore 1 - \beta = 1 - (1 - c^{\theta_1})$$

$\therefore$ Power of the test, $\boxed{\therefore 1-\beta = c^{\theta_1}}$

LIKELIHOOD RATIO TEST (LRT)

NP lemma based on the ratio of the two ~~probabilities~~ probability density functions provides the best test for simple null hypothesis against simple alternative hypothesis. In this case, the best depends

on the nature of the population distribution and the form of the alternative hypothesis considered. But, likelihood ratio test suggest a sample method of test construction which is closely related to MLE (Maximum Likelihood Estimator), hence the likelihood ratio (LR) gives the method of testing composite hypothesis.

① Likelihood ratio test for mean of a normal population when variance is known.

z-test	t-test
$n > 30$ or popⁿ s.d. σ is known	$n \leq 30$ and popⁿ s.d. σ is unknown

variance is known $\Rightarrow$ z-test
variance is unknown $\Rightarrow$ t-test
$n > 30$ (z-test)
$n \leq 30$ (t-test)

<u>Procedure:</u>

$H_0: \mu = \mu_0$ i.e. there is no significant difference between sample mean and population mean.
$H_1: \mu \neq \mu_0$ i.e. there is a significant difference between sample mean and population mean. (Two-Tailed Test)

or, $\left.\begin{array}{l} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{array}\right\}$ (One Tailed Test)

<u>Test-statistic:</u> Under H_0,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

<u>Decision:</u> If $|z_{cal}| > |z_{tab}|$ then we reject H_0 otherwise do not reject H_0.

[<u>NOTE</u>: If α is not given, consider it as 5%.]

(ii) LRT for mean of a normal population when variance is unknown.

<u>Test-statistic</u>: Under H_0,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where, s = sample s.d. $= \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2 - n\bar{x}^2}{n-1}}$ (For making estimate unbiased)

<u>Critical Value</u>: $t_{tab} = t_{n-1,\,\alpha}$

where, $n-1$ = degree of freedom.

<u>Decision</u>: If $|t_{cal}| \geq |t_{tab}|$, then we reject H_0, otherwise do not reject H_0.

[3 marks]

Q1. Given. $\mu = 400$,
$\sigma = 20$,
$n = 100$,
$\bar{x} = 390$
$\alpha = 5\% = 0.05$

Notation
n = Sample size
N = Population size
σ = Pop^n s.d.
s = sample s.d.
μ = Pop^n mean
$\bar{x}$ = Sample mean

H_0: $\mu = 400$ i.e. the claim is valid / correct.
H_1: $\mu \neq 400$ i.e. the claim is incorrect.

<u>Test-Statistic</u>: Under H_0,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{390-400}{20/\sqrt{100}} = \frac{-10(10)}{20}$$

$$\Rightarrow z_{cal} = -5$$

Scanned with CamScanner

Critical value:

We have, $\alpha = 0.05$

$\therefore Z_{tab} = Z_{0.05}$ (Two tailed test) $= 1.96$

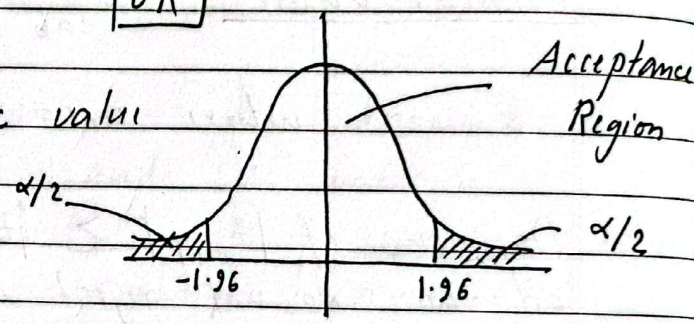
$-3.99 < Z < 3.99$

Decision: Since $|Z_{cal}| > Z_{tab} = 1.96$, so we reject H_0

Hence, the claim is incorrect.

OR

Since, the test-statistic value lies in the rejection region, so we reject H_0.



Q3. Given,

$\mu = 75$

$n = 25$

$\bar{x} = 83.40$

$s = 11.83$

$\alpha = 0.05$

$H_0: \mu = 75$ i.e. the average balance is not more than \$75.

$H_1: \mu > 75$ i.e. the average balance is more than \$75. (Right or one-tailed test)

Test-statistic: Under H_0,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{83.40 - 75}{11.83/\sqrt{25}}$$

Scanned with CamScanner

$= 3.55$

Critical value:

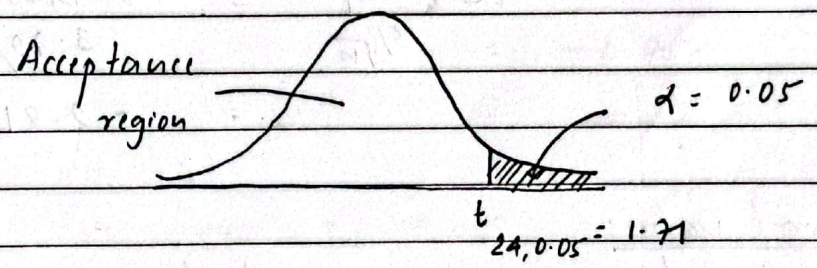
We have, $\alpha = 0.05$ and $dof = n-1 = 25-1 = 24$

$\therefore t_{tab} = t_{24,\,0.05}$ (One Tailed Test)

$= 1.71$

Decision: Since, $t_{cal} > t_{tab}$, so we reject H_0.
Hence, the ~~Avg~~ average monthly balance is more than \$75.

OR



Since, the test-statistic value lies in the rejection region so we reject H_0.

Q.4

Let x = length of stay (in days).

$\Sigma x = 10+15+11+5+7+4+8+14+10 = 85$ and $\Sigma x^2 = 100+225+121+25+49+16+64+196+121 = 917$

$\therefore n = 9$

$\bar{x} = \frac{\Sigma x}{n} = \frac{85}{9} = \cancel{9.44}\ \cancel{9.55}\ 9.44$

$y = 13$

$$s.d.(s) = \sqrt{\frac{\Sigma x^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{917 - 9(9.44)^2}{9-1}}$$

$$= \underline{3.79}$$

$H_0: \mu = 13$ i.e. the average length of stay of tourist is 13 days

$H_1: \mu \neq 13$ i.e. the average length of stay of tourist is not 13 days (Two-Tailed Test)

<u>Test-statistic</u>: Under H_0,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{9.44 - 13}{3.79/\sqrt{9}}$$

$$= -2.81.$$

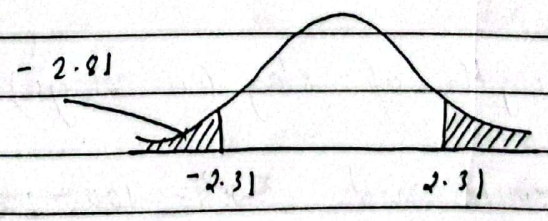
<u>Critical value:</u>

We have, $\alpha = 0.05$ and $dof = n-1 = 9-1 = 8$

$$\therefore t_{tab} = t_{8, 0.05} \text{ (Two-Tailed Test)}$$

$$= 2.31$$

<u>Decision</u>



Since, test-statistic value lies in the rejection region so we reject H_0.

Hence, the claim is incorrect.

Scanned with CamScanner

Q15

Given, $\mu = 2.8$, $\sigma = 0.20$

$n = 25$, $\bar{x} = 2.7$

$\alpha = 0.05$

H_0: $\mu \geq 2.8$ i.e. the production equipment does not need to be adjusted.

H_1: $\mu < 2.8$ i.e. the production equipment needs to be adjusted. (Left-Tailed Test)

Test-statistic: Under H_0,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.73 - 2.8}{0.20/\sqrt{25}}$$

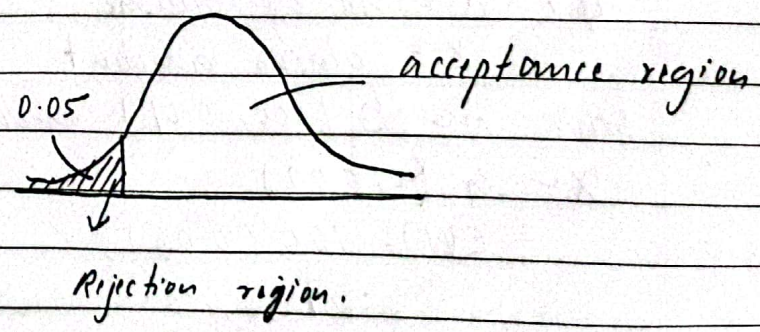
$$= -1.75$$

Critical value:

We have, $\alpha = 0.05$

$\therefore Z_{tab} = Z_{0.05}$ (Left-tailed Test) $= -1.645$

Decision:



Since, test-statistic value lies in the rejection region, so we reject H_0. Hence, the production equipment needs to be adjusted.

Assignment: Q17, Q16, Q9, Q6

Scanned with CamScanner

[LG-1 SG-1]

Unit 5:
Compound and mixed type distribution.

Introduction

A random variable $S = X_1 + X_2 + \dots + X_N$ is said to have a compound distribution if,

(i) N is a random variable

(ii) X_is are i.i.d. (identically, independently distributed)

(iii) Each X_is ~~are~~ is independent of N.

Mean and Variance of 's'

$E(S) = E(X) \cdot E(N)$

$V(S) = E(N) \cdot V(X) + V(N) [E(X)]^2$

Q Suppose number of claims has a Binomial $(100, 0.01)$ distribution and claim sizes have a Gamma $(10, 0.02)$ distribution. Find the mean and variance of aggregate claim amount 's'.

Solution

Let, N = no. of claims

X = claim amount

Since $N \sim B(100, 0.01)$ and $X \sim G(10, 0.02)$

$\therefore E(N) = 100(0.01)$

$= 1$

$\therefore V(N) = 100(0.01)(1-0.01)$

$= 0.99$

$X \sim B(n, p)$

$\Rightarrow E(X) = np$

$V(X) = npq$

$X \sim G(\alpha, \beta)$

$E(X) = \alpha/\beta$

$V(X) = \alpha/\beta^2$

$E(X) = 10/0.02 = 500$

& $V(X) = \dfrac{10}{(0.02)^2} = \dfrac{500}{0.02} = 25000$

Now, the mean and variance of aggregate claim amount.
S

$E(S) = E(N) \cdot E(X) =$

$V(S) = E(N)\,V(X) + V(N)\,[E(X)]^2$

$=$

H/W If X has an exponential distribution with parameter $\theta = 10$ and N has a Poisson with $\lambda = 50$ distribution, calculate the expected value and variance of 'S'.

$X \sim Exp(\theta)$

$E(X) = \frac{1}{\theta}$ $\quad V(X) = \frac{1}{\theta^2}$

$X \sim P(\lambda)$

$E(X) = V(X) = \lambda$

Nov 16, 2024

[Compound Negative Exponential Distribution]

Let us assume that N follows Poisson distribution with parameter λ, and each X_i follows negative exponential distribution with parameter μ. The compound distribution ~~with~~ Y is the sum of N independent and identically distributed negative exponential random variables.

Poisson Distribution

Exponential Dist^n

Here, $N \sim P(\lambda) \Rightarrow P(N=n) = \dfrac{e^{-\lambda}\lambda^n}{n!}$

and $X_i \sim Exp(\mu) \Rightarrow f(x_i) = \mu e^{-\mu x_i}$; $x_i \geq 0$

Now, the compound distribution Y is given by the convolution of dist^n of N and x_is

$$f_Y(y) = \sum_{n=0}^{\infty} P(N=n)\, f_{x_1+x_2+\dots+x_n}(y)$$

Assuming independence, the sum of N negative exponential random variables follows a Gamma distribution with the shape parameter 'n' and rate parameter 'μ'.

$$\therefore f_{x_1+x_2+\dots+x_n}(y) = \frac{\mu^n y^{n-1} e^{-\mu y}}{(n-1)!} \qquad \begin{pmatrix} \alpha = n \\ \beta = \mu \end{pmatrix}$$

$\therefore$ The pdf of compound negative exponential distribution is given by

$$f_Y(y) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \frac{\mu^n y^{n-1} e^{-\mu y}}{(n-1)!}$$

Date
Page

Mean and Variance:

Let N be a Poisson distribution with parameter λ representing the number of events and let $X_1, X_2, \ldots, X_N$ be independently and identically distributed random variables following exponential distn with the parameter μ.

Then the mean and variance of compound negative exponential distribution is given by,

$$E(Y) = E(N)\,E(X) \quad [Y = X_1 + X_2 + \ldots + X_N]$$
$$= \lambda \cdot \frac{1}{\mu}$$
$$= \frac{\lambda}{\mu}$$

and
$$V(Y) = E(N)\,V(X) + V(N)\,[E(X)]^2$$
$$= \lambda \frac{1}{\mu^2} + \lambda \left[\frac{1}{\mu}\right]^2$$
$$= \frac{\lambda}{\mu^2} + \frac{\lambda}{\mu^2}$$
$$\boxed{\therefore V(Y) = \frac{2\lambda}{\mu^2}}$$

[6M] Sure

Mixed distribution:

A mixed random variable is a type of random variable that follows a mixture distribution. A mixture distribution is a probability distribution derived from a combination of two or more component distributions.

$f(x)$ = prob. density function (pdf)

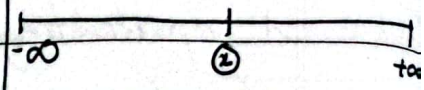
$F(x)$ = Cumulative distribution function (cdf).

Q. The probability dist^n function of a random variable X is

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2-x & \text{for } 1 \le x < 2 \\ 0 & \text{for } x \ge 2 \end{cases}$$

Compute the cumulative distribution of X.

$= P(X \le x)$



$= \sum_{x} P(x)$ [Discrete

$= \int_{-\infty}^{x} f(x)\,dx$ [Continuous

$f(x) = \frac{dF(x)}{dx}$

Solution

For any x in the range $x \le 0$, we have:

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} 0\,dx = 0$$

For x in the range $0 < x < 1$, we have

$$F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx$$

$$= \int_{-\infty}^{0} 0\,dx + \int_{0}^{x} x\,dx$$

$$\boxed{F(x) = \frac{x^2}{2}}$$

For x in the range $1 \le x < 2$, we have

$$F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{x} f(x)\,dx$$

$$= 0 + \int_{0}^{1} x\,dx + \int_{1}^{x} (2-x)\,dx$$

$$= 0 + \frac{1}{2} + \left[2x - \frac{x^2}{2}\right]_1^x$$

$$= \frac{1}{2} + \left(2x - \frac{x^2}{2} - 2 + \frac{1}{2}\right)$$

$$= \frac{1}{2} + \left(2x - \frac{x^2}{2} - \frac{3}{2}\right)$$

$$F(X) = \left(2x - \frac{x^2}{2} - 1\right)$$

For x in the range $x > 2$

$$F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$$

$$= 0 + \frac{1}{2} + \int_{1}^{2} (2-x)\,dx + \int_{2}^{\infty} 0\,dx$$

$$= 0 + \frac{1}{2} + \left[2x - \frac{x^2}{2}\right]_1^2 + 0$$

$$= \frac{1}{2} + \left(4 - 2 - 2 + \frac{1}{2}\right)$$

$$= 1 + 0$$

$$= \boxed{1}$$

Thus, we have,

$$F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x^2}{2} & 0 < x < 1 \\ 2x - \frac{x^2}{2} - 1 & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

H/W Find the distribution function of random variable X whose density function is given by

$$f(x) = \frac{x}{2} \quad \text{for } 0 < x < 1$$

$$= 1/2 \quad \text{for } 1 \le x < 2$$

$$= \frac{3-x}{2} \quad \text{for } 2 \le x < 3$$

$$= 0 \quad \text{for } x \ge 3$$

Q The time taken to perform a particular task, x hrs, has the probability density function

$$f(x) = \begin{cases} 10\,cx^2 & ; \ 0 \le x < 0.6 \\ 9c(1-x) & ; \ 0.6 < x < 1.0 \\ 0 & ; \ \text{otherwise} \end{cases}$$

where 'c' is a constant

(a) Find the value of 'c'

(b) Find the expected time

(c) Determine the prob. that the time will be

(i) more than 0.8 hrs

(ii) betn 0.4 and 0.8 ~~minutes~~. hours

$$\int_x f(x)\,dx = 1$$

Solution

(a) $\because \int_x f(x)\,dx = 1$

$$\Rightarrow \int_0^1 f(x)\,dx = 1$$

$$\Rightarrow \int_0^{0.6} f(x)\,dx + \int_{0.6}^{1} f(x)\,dx = 1$$

$$\Rightarrow 10c\int_0^{0.6} x^2\,dx + 9c\int_{0.6}^{1} (1-x)\,dx = 1$$

$$\Rightarrow 10c\left[\frac{x^3}{3}\right]_0^{0.6} + 9c\left[x - \frac{x^2}{2}\right]_{0.6}^{1} = 1$$

$$\Rightarrow 10c\left(\frac{0.6^3}{3}\right) + 9c\left[1/2 - 0.6 + \frac{(0.6)^2}{2}\right] = 1$$

$$\Rightarrow 0.72c + 0.72c = 1$$

$$\therefore \boxed{c = \frac{1}{1.44}} = \underline{0.6944}$$

$$\therefore f(x) = \begin{cases} 10/1.44\, x^2; & 0 \le x < 0.6 \\ \frac{9}{1.44}(1-x); & 0.6 \le x < 1.0 \\ 0; & \text{otherwise} \end{cases}$$

(b)

$$E(X) = \int_x x\, f(x)\,dx$$

$$= \frac{10}{1.44}\int_0^{0.6} x^3\,dx + \frac{9}{1.44}\int_{0.6}^{1} x(1-x)\,dx$$

$$= \frac{10}{1.44}\left[\frac{x^4}{4}\right]_0^{0.6} + \frac{9}{1.44}\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{0.6}^{1}$$

$$= \frac{10}{1.44}\left(\frac{0.6^4}{4}\right) + \frac{9}{1.44}\left[\frac{1}{2} - \frac{1}{3} - \frac{(0.6)^2}{2} + \frac{(0.6)^3}{3}\right]$$

$$= 0.225 + 6.25 * (-0.01 + 0.072)$$

$$= 0.225 + 0.3875$$

$$= 0.612$$

$$= \boxed{0.592}$$

(c) (i) $P(X > 0.8) = \int_{0.8}^{1} f(x)\,dx = \frac{9}{1.44}\int_{0.8}^{1} (1-x)\,dx$

$= \frac{9}{1.44}\left[x - \frac{x^2}{2}\right]_{0.8}^{1} = \frac{9}{1.44}\left[1 - \frac{1}{2} - 0.8 + \frac{0.8^2}{2}\right]$

$= \boxed{0.125}$

(ii) $P(0.4 < x < 0.8) = \int_{0.4}^{0.6} f(x)\,dx + \int_{0.6}^{0.8} f(x)\,dx$

$= \frac{10}{1.44}\int_{0.4}^{0.6} x^2\,dx + \frac{9}{1.44}\int_{0.6}^{0.8} (1-x)\,dx$

$= \frac{10}{1.44}\left[\frac{x^3}{3}\right]_{0.4}^{0.6} + \frac{9}{1.44}\left[x - \frac{x^2}{2}\right]_{0.6}^{0.8}$

$= \frac{10}{1.44}\left(\frac{0.6^3 - 0.4^3}{3}\right) + \frac{9}{1.44}\left[0.8 - \frac{(0.8)^2}{2} - 0.6 + \frac{(0.6)^2}{2}\right]$

$= 0.352 + 0.375$

$= \boxed{0.727}$

Q. Let, X have a mixed distribution with F(x) defined as follows:

$$F(x) = \begin{cases} 0 & \text{if } x < -2 \\ \frac{1}{12}(2+x) & \text{if } -2 \le x < 0 \\ \frac{1}{6}(1+x) & \text{if } 0 \le x < 4 \\ \frac{1}{12}(6+x) & \text{if } 4 \le x < 6 \\ 1 & \text{if } x \ge 6 \end{cases}$$

obtain the mean and variance of X.

Date
Page

Solution

The pdf does not exist for $x < -2$ and $x \geq 6$.

Since, $f(x) = \dfrac{dF(x)}{dx}$

For $-2 \leq x < 0$, $f(x) = \dfrac{d}{dx}\left(\frac{1}{12}(2+x)\right) = \frac{1}{12}$

$0 \leq x < 4$, $f(x) = \dfrac{d}{dx}\left(\frac{1}{6}(1+x)\right) = \frac{1}{6}$

$4 \leq x < 6$, $f(x) = \dfrac{d}{dx}\left(\frac{1}{12}(6+x)\right) = \frac{1}{12}$

Mean:

$$E(X) = \int_x x\, f(x)\, dx$$

$$= \int_{-2}^{0} x\, f(x)\, dx + \int_{0}^{4} x\, f(x)\, dx + \int_{4}^{6} x\, f(x)\, dx$$

$$= \int_{-2}^{0} \frac{1}{12} x\, dx + \int_{0}^{4} \frac{1}{6} x\, dx + \int_{4}^{6} \frac{1}{12} x\, dx$$

$$= \left[\frac{x^2}{24}\right]_{-2}^{0} + \left[\frac{x^2}{12}\right]_{0}^{4} + \left[\frac{x^2}{24}\right]_{4}^{6}$$

$$= -\frac{1}{6} + \frac{16}{12} + \frac{36}{24} - \frac{16}{24}$$

$$= \frac{-8 + 64 + 72 - 32}{48}$$

$$= \frac{104 - 8}{48} = \frac{96}{48} = \boxed{2}$$

Variance

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

Now, $E(X^2) = \int_x x^2 f(x)\, dx$

$$= \int_{-2}^{0} x^2 f(x)\, dx + \int_{0}^{4} x^2 f(x)\, dx + \int_{4}^{6} x^2 f(x)\, dx$$

$$= \int_{-2}^{0} \frac{1}{12} x^2\, dx + \int_{0}^{4} \frac{1}{6} x^2\, dx + \int_{4}^{6} \frac{1}{12} x^2\, dx$$

$$= \left[\frac{x^3}{36}\right]_{-2}^{0} + \left[\frac{x^3}{18}\right]_{0}^{4} + \left[\frac{x^3}{36}\right]_{4}^{6}$$

$$= \frac{8}{36} + \frac{64}{18} + \frac{\cancel{196}\ 216}{36} - \frac{64}{36}$$

$$= \frac{8 + 128 + \cancel{196}\ 216 - 64}{36}$$

$$= 7.44$$

$9.77 - 2^2 =$

[check] $\therefore V(X) = 7.44 - (2)^2 = ____ - 4 = ____$

Extreme Value Distribution (EVD)

EVD is a statistical distribution used to model the extreme values (either maximum or minimum) in a set of observations. It is particularly relevant in fields of insurance, hydrology, finance and environmental science, where the focus is on ~~extreme science~~ the extreme events.

If we are dealing with losses that have typical sizes, i.e. ones whose values come from the central part of the distⁿ, we can make use of central limit theorem. This tells us that, if we calculate the mean, $\bar{x}$ of a set of 'n' values ~~that~~ taken from a loss distribution ~~of~~ that has mean μ and variance σ^2, the standardized value, $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, can be approximated using the standard normal distribution.

However, the most financial part of a loss distⁿ is usually the right-hand tail where the large losses occur. These are the extreme values of distⁿ. So, there is a similar way to approximate the behavior of the extreme values in the tail of the distⁿ.

The Generalized Extreme Value (GEV) distⁿ is a family of distⁿ with the EVD.

The pdf of GEV distⁿ is given by:

$$f(x) = \frac{1}{\beta}\left[1+\frac{\gamma(x-\alpha)}{\beta}\right]^{-(1+1/\gamma)} \exp\left[-\left(1+\frac{\gamma(x-\alpha)}{\beta}\right)^{1/\gamma}\right]$$

where, X = the random variable representing extreme values

γ = The shape parameter, determining the type of distribution

$\alpha =$ a location parameter
$\beta =$ a scale parameter.

Applications of EVD

(i) EVD is widely used in hydrology to model the distribution of extreme values of river flows, rainfall and flood peaks. This is crucial for design infrastructure such as dams, bridges etc.

(ii) In finance, extreme value analysis is used for modeling extreme returns and assessing the risk associated with extreme market events.

(iii) In the insurance industry, extreme value distributions are used to model the tail risk associated with rare and severe events.

This is crucial setting insurance premiums and reserves to cover potential extreme losses.

(iv) In the field of telecommunications, extreme value analysis is applied to model extreme traffic loads on communication network. This helps in designing network. This helps in designing network that can handle rare but intense periods of high demand.