11/12/2023
Statistics

Md. Mehedi Hasan Rafy

Section – A

1. sampling distribution
   - χ^2 (Chi-square)
   - t distribution
   - F-distribution
   - Z distribution/Normal distribution

2. Estimation
   - point
   - properties

Section – B :

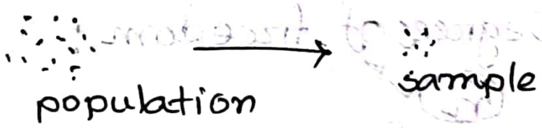
contingency table
Test of significance
Non parametric

Standard Normal Z — In which normal distribution Mean = 0, varience = 1.

$$Z = \frac{X-\mu}{\sqrt{V(X)}} = \frac{X-\mu}{\sigma}$$

$$z(\mu, \sigma^2) = (0, 1)$$

population:



Random variable:

*parameter is a constant that species population distributi

Ex:

$X \sim N(\mu, \sigma^2)$, μ and σ^2 are parameters of Normal d.

$p(X=x) = p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$, λ is the parameter of the binomial distribution.

$$\bar{x} = \frac{\sum x_i}{N} = \frac{x_1 + x_2 + \cdots x_n}{n}$$

This is the function of sample observation.

↳ statistic

***What is statistic?

— function of sample observation.

Example: $\bar{x}$, s^2, r

probability distribution of population:

***Sampling distribution:

(i) standard normal

(ii) chi-square

(iii) T

(iv) F

(I) Definition
(II) Derivation
(III) Equation
(iv) Mean
(v) Variance
(vi) Skewness β_1
(vii) Kurtosis, β_2
(viii) Moment Generating Function
(ix) properties
(X) Application
(xi) Usefulness

- For Exam

$z^2 \longrightarrow \chi^2$ distribution

$$z = \frac{x - \mu}{\sigma}$$

↳ Range: $-\infty$ to ∞

Range (chi-square): 0 to ∞

Degrees of freedom:

$\sum_{i=1}^{n} z_i^2 \longrightarrow$ degrees of freedom $= n$

consists of ... (set of elements

(i) V = vertices, nodes, points (set of elements

(ii) E = edges identified with a unique (unordered pairs) $[u, v]$.

~~Ass~~

12/12/2023

Statistics

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

X is a random variable which follows a normal distribution which has a mean of μ and ~~standard deviation~~ variance σ^2.

Probability Density Function

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

$= 0$ otherwise

range:

$-\infty < x < \infty$

$-\infty < \mu < \infty$

$\sigma^2 > 0$

properties

(I) $0 \leq f(x) \leq 1$

(II) $\sum f(x) = 1$ as it is a continuous distribution

we can write,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

Derivation of χ^2

Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z = \frac{x-\mu}{\sigma} \sim N(0,1) \longrightarrow \text{Standard normal.}$$

$$\Rightarrow z^2 = \left(\frac{x-\mu}{\sigma}\right)^2 \sim \chi^2 \text{ (chi-square) with d.f} = 1.$$

$$\Rightarrow \sum_{i=1}^{n} z^2 = \sum_{i=1}^{n} \left(\frac{x_i-\mu}{\sigma}\right)^2 \sim \chi^2 \text{ with } n \text{ d.f.}$$

Chi-square

A chi-square (χ^2) distribution is a continuous probability distribution that is used in many hypotheses tests. The chi-square distribution may be defined as the "sum of the squares of independent normally distributed variables with zero means and unit variances".

It is normally pronounced as ki-square with 1 degrees of freedom.

CamScanner

pdf of chi square

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{n}{2}-1} \quad , \chi^2 \geq 0$$

$$= 0 \text{ , otherwise}$$

if ~~f(x)~~ $x \sim \chi^2_{(n)}$ then pdf $f(x)$ is denoted as,

$$f(x) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{\frac{x}{2}} (x)^{\frac{n}{2}-1}$$

Mean ~~of~~ $E(X) = \int_0^\infty x f(x) dx$

Variance $= V(X) = E(x^2) - \{E(x)\}^2$

$E(x^2) = \int_0^\infty x^2 f(x)\, dx$

Mean of chi-square distribution

*** • Find the moment generating function of chi-square distribution, Hence or otherwise find mean, variance, β_1, β_2 of χ^2 distribution

Moment Generating Function (mgf) : By the definitio

$$M_x(t) = E(e^{tx}) = \int_x e^{tx} f(x)\, dx$$

$$= \int_0^\infty e^{tx} \frac{1}{2^{n/2} \gamma \frac{n}{2}} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} dx$$

$$= \frac{1}{2^{n/2} \gamma \frac{n}{2}} \cdot \int_0^\infty e^{tx - \frac{x}{2}} x^{\frac{n}{2}-1} dx$$

$$M_x(t) = (1-2t)^{-\frac{n}{2}} \quad |\, |2t| < 1 .$$

12/12/2023

Accounting

18/12/2023

Statistics

Moment Generating Function (mgf)

$x \sim \chi_n^2$ then, $M_x(t) = (1-2t)^{-\frac{n}{2}}$

$2t<1$

$\therefore t<\frac{1}{2}$

$$\boxed{\mu_r' = \frac{\delta^r}{\delta t^r} M_x(t) \quad \text{After derivation put } t=0}$$

$\mu_1' = \frac{\delta}{\delta t} M_x(t)$

$= \frac{\delta}{\delta t}\{(1-2t)^{-n/2}\}$

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

$= n(1-2t)^{-\left(\frac{n+2}{2}\right)}$

$= n(1-2.0)^{-\left(\frac{n+2}{2}\right)} \quad | t=0$

$= n,$

Mean $= \mu_1 = \mu_1' = n$

$\mu_2' = \cancel{n-2nt}\; n(1-2t)^{-\left(\frac{n+2}{2}\right)}$

$= -n\cdot\left(\frac{n+2}{\cancel{2}}\right)(1-2t)^{-\left(\frac{n+4}{2}\right)}(\cancel{-2})$

$= n^2+2n \quad | t=0$

$= n(n+2)$

Variance, $\mu_2 = \mu_2' - \mu_1'^2$

$$= n(n+2) - n^2$$

$$= n^2 + 2n - n^2$$

$$= 2n.$$

* Show that variance of χ^2 distribution is twice of its mean.

<u>Cumulant Generating Function:</u>

$$K_X(t) = \log M_X(t) = \log (1-2t)^{-\frac{n}{2}} = -\frac{n}{2}\log(1-2t)$$

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots\cdots \frac{x^n}{n}. \qquad = \frac{n}{2}\left(2t + \frac{(2t)^2}{2} + \frac{(2t)^3}{3} + \cdots\right)$$

κ_r = coefficient of $\frac{t^r}{r!}$ in $K_X(t)$

κ_1 = coefficient of $\frac{t}{1!}$ in $K_X(t)$

$$= \frac{n}{2}\cdot 2\cancel{t}$$

$$= n.$$

κ_2 = coefficient of $\frac{t^2}{2!}$ in $K_X(t)$

$$= \frac{n}{2}\cdot 4\left(\frac{t^2}{2}\right)$$

$$= 2n.$$

κ_3 = coefficient of $\frac{t^3}{3!}$ in $K_X(t)$

$$= \frac{n}{2}\cdot\frac{8t^3}{3} = \frac{8nt^3}{3!}$$

$$= 8n$$

κ_4 = coefficient of $\frac{t^4}{4!}$ in $K_x(t)$

$$= \frac{n}{2}\cdot\frac{(2t)^4}{4}$$

$$= \frac{n}{2}\cdot\frac{16t^4}{4}$$

$$= \frac{48nt^4}{24}$$

$$= \frac{48nt^4}{4!}$$

$$= 48n.$$

$\mu_1 = \kappa_1 = \text{Mean} = n$

$\mu_2 = \kappa_2 = \text{Variance} = 2n$

$\mu_3 = 8\kappa_3 = 8n$

$\mu_4 = \kappa_4 + 3\kappa_2^2$

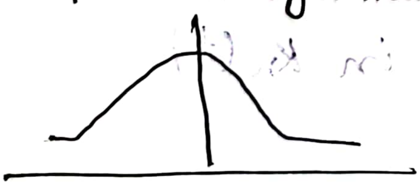
$= 48n + 3(2n)^2$

$= 12n^2 + 48n.$

<u>Coefficient of skewness:</u>

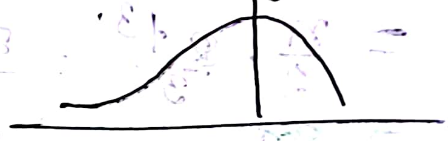
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(8n)^2}{(2n)^3} = \frac{64n^2}{8n^3} = \frac{8}{n}$$

$\beta_1 = 0$ = symmetric

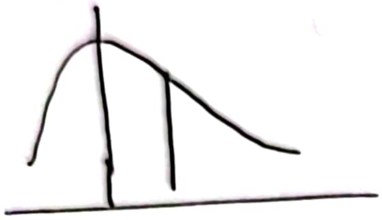


$\beta_1 \neq 0$ asymmetric

$\mu_3 < 0$ negative skewness



$\mu_3 > 0$ positive skewness



Coefficient of Kurtosis:

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2} = \frac{12n(n+4)}{(2n)^2} = \frac{\cancel{12}^{3}\cancel{n}(n+4)}{\cancel{4}\cancel{n^2}n} = \cancel{3+\frac{12}{n}}$$

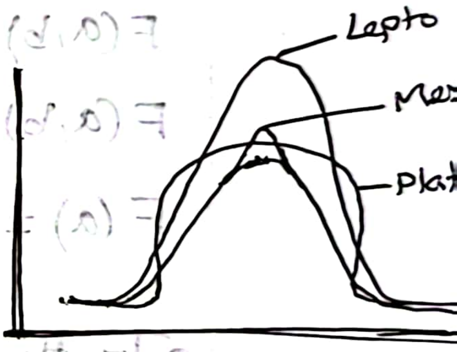
$$= 3 + \frac{12}{n}.$$

if,

$\beta_2 = 3 \longrightarrow$ Mesokurtic

$\beta_2 < 3 \longrightarrow$ Leptokurtic

$\beta_2 > 3 \longrightarrow$ Platykurtic.



- Applications of chi-square distribution.

- – –

- Limiting case:

if $n \to \infty$ $(n > 30)$

then, $\chi^2_{n\to\infty} \longrightarrow$ Normal distribution.

$\cancel{x_{t_n}}$. x_1 and x_2 are indepent and $x_1 \to \chi^2_{n_1}$ and $x_2 \to \chi^2_{n_2}$ then $x_1 \pm x_2 \to \chi^2_{n_1 \pm n_2}$.

$$= e^{\frac{\mu t}{\sigma}}(1) M_{\frac{x}{\sigma}}(t)$$

$$= e^{\frac{\mu t}{\sigma}} \cdot M_x\left(\frac{t}{\sigma}\right)$$

$$= e^{\frac{\mu t}{\sigma}}\left(1 - \frac{2t}{\sigma}\right)^{-\frac{n}{2}}$$

$$= e^{\frac{nt}{\sqrt{2n}}}\left(1 - \frac{2t}{\sqrt{2n}}\right)^{-\frac{n}{2}}$$

$$= e^{t\sqrt{\frac{n}{2}}}\left(1 - t\sqrt{\frac{2}{n}}\right)^{-\frac{n}{2}}$$

Mean $= \mu = n$

variance $= \sigma^2 = 2n$

$\sigma = \sqrt{2n}$

$$M_z(t) = e^{t\sqrt{\frac{n}{2}}}\left(1 - t\sqrt{\frac{2}{n}}\right)^{-\frac{n}{2}}$$

$$K_z(t) = \log M_z(t)$$

$$= -t\sqrt{\frac{n}{2}} - \frac{n}{2}\log\left(1 - t\sqrt{\frac{2}{n}}\right)$$

$$= -t\sqrt{\frac{n}{2}} + \frac{n}{2}\left(t \cdot \sqrt{\frac{2}{n}} + \frac{t^2}{2} \cdot \frac{2}{n} + \frac{t^3}{3}\left(\frac{2}{n}\right)^{\frac{3}{2}} + \dots\right]$$

$$= -t\sqrt{\frac{n}{2}} + t \cdot \sqrt{\frac{n}{2}} + \frac{t^2}{2} + O\left(n^{-\frac{1}{2}}\right)$$

$$= \frac{t^2}{2} + O\left(n^{-1/2}\right)$$

Where $O(n^{-\frac{1}{2}})$ are terms containing $n^{\frac{1}{2}}$ and higher powers of n in the denominator.

$$\lim_{n\to\infty} K_z(t) = \frac{t^2}{2} \Rightarrow M_z(t) = e^{\frac{t^2}{2}} \text{ as } n \to \infty$$

which is the mgf of standard normal variate.

Additive property of χ^2

Theorem:

Sum of independent χ^2-variate is also χ^2-variate.
Suppose X_i $(i=1,2,\dots,K)$ are K-independent χ^2 variate with n_i $(i=1,2,\dots,K)$ degrees of freedom respectively, then,

$\sum_{i=1}^{K} X_i$ follows χ^2-variate with $n = \sum_{i=1}^{K} n_i$ d.f.

proof:

Since $X_i \sim \chi^2$-variate with n_i $(i=1,2,\dots,K)$ d.f.
then by definition of mgf we have,

$$M_{x_i}(t) = (1-2t)^{-\frac{n_i}{2}} \quad (i=1,2,\dots,K)$$

$$M_{x_i}(t) = \cancel{M_{x_1}t + M_{x_2}t +} = M_{x_1+x_2+x_3+\dots x_K}(t)$$

$$= M_{x_1}(t)\, M_{x_2}(t) \dots\dots M_{x_K}(t)$$

$$= (1-2t)^{-\frac{n_1}{2}} (1-2t)^{-\frac{n_2}{2}} \dots (1-2t)^{-\frac{n_K}{2}}$$

$$= (1-2t)^{-\frac{1}{2}(n_1+n_2+\dots n_K)}$$

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

which is the mgf of χ^2-variate with $n = \sum_{i=1}^{K} n_i$ d.f.
Hence by the uniqueness theorem of mgf is necessary.

Mode of χ^2 – variate

The mode of χ^2- variate is the maximum of χ^2 func

We know the pdf,

step

① $f(x) = \dfrac{1}{2^{\frac{n}{2}}\Gamma\frac{n}{2}} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} \; ; x \geq 0$

② $f'(x) = 0$

$$\log f(x) = \log(c) - \frac{x}{2} + \left(\frac{n}{2}-1\right)\log x$$

$$\Rightarrow \cancel{\log}\frac{f'(x)}{f(x)} = 0 - \frac{1}{2} + \left(\frac{n}{2}-1\right)\frac{1}{x}$$

Now,

$$\cancel{\Rightarrow} \log \frac{f'(x)}{f(x)} = 0$$

Since $f(x) \neq 0$ then $f'(x) = 0$.

$$\cancel{f'(x) =} \therefore -\frac{1}{2} + \left(\frac{n}{2}-1\right)\frac{1}{x} = 0$$

$$\Rightarrow \left(\frac{n}{2}-1\right)\frac{1}{x} = \frac{1}{2}$$

$\frac{n-2}{2}$

$$\Rightarrow \frac{1}{x} = \frac{1}{\cancel{2}} \times \frac{\cancel{2}}{n-2} = \frac{1}{n-2}.$$

$$\therefore x = n-2.$$

$(\bar{p}+r)(\bar{q}+r)$

$\cancel{\bar{p}\bar{q}+}\boxed{\bar{p}+\bar{q}}+r$

$\overline{pq}+r$

$(\bar{p}+q)(\bar{q}+p) = \neg(p \leftrightarrow q) \vee r$

$= \bar{p}\bar{q} + \bar{p}p + q\bar{q} + pq$

$= \bar{p}\bar{q} + pq$

$$f''(x) = 0 - \left(\frac{n}{2}-1\right)\frac{1}{x^2}$$

$$= \frac{1}{x^2}\left(1-\frac{n}{2}\right) = \frac{1}{n^2}\left(\frac{2-n}{2}\right) = -\frac{1}{x^2}\left(\frac{n-2}{2}\right)$$

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

$$= -\frac{1}{2(n-2)} < 0, \text{ for } n$$

22/01/24

Statistics

Sampling T dis'n

- student's: $\nu = n-1$ d.f.
- fisher's: n d.f.

*** **Q.** Define T dis'n with its major application?

Ans

Student's t distribution is a sampling distribution of continuous types range's from $-\infty$ to ∞.

Derivation:

Let $x_1, x_2, \ldots x_n$ be a random sample of size n from a normal population with mean μ and variance σ^2. then students t is defined by the statistics

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}.$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean and $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$ is an unbiased estimate of the population variance, σ^2, and it follows students t dis'n. with $\nu = (n-1)$ degree of freedom with probability density function

$$f(t) = \frac{1}{\sqrt{\nu}\, \beta\left(\frac{1}{2}, \frac{\nu}{2}\right)\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}, \quad -\infty < t < \infty$$

, 0 otherwise

Derivations of student's t dis'n

$$\boxed{\begin{aligned} &\text{if } \quad n = (n-1) = 1 \\ &f(t) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2},\frac{1}{2}\right)} \cdot \frac{1}{(1+t^2)} \end{aligned}} \quad -\infty < t < \infty$$

which is standard Chauchy Distribution.

Derivation

fishers t

$t = \dfrac{z}{\sqrt{\frac{x}{n}}}$ where (z follows standard norma

$z \sim (0,1)$ and $x \sim \chi^2$ with n d.f.

It is the ratio of a standard normal variate to the square root of an independent χ^2 variate divided by its degrees of freedom

if z is a standard normal variate(0,1) and x is an independent χ^2 - variate with n degree of freedom then fisher's T is given by $t = \frac{z}{\sqrt{\frac{x}{n}}}$. and it follows students t with n degrees of freedom an pdf is given by,

$$f(t) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{t^2}{2}\right)^{\frac{n+1}{2}}}, \quad -\infty < t < \infty$$

$= 0$, otherwise

if $n > 30$ $f(t)$ becomes normal distn

Application

(1) It is used to test the single sample mean, $\bar{x}$. with a population mean μ.

(2) To compare the population means. we can use t-dis'n.

(3) Used to test correlation and regression co-efficient.

(4) To test for proprotion test.

(5) ~~To~~ test for multiple correlation coefficien

Derivation of fisher's T

if x, y are independent

$$f(x,y) = f(x)f(y)$$

$$f(x,y) = x^{\sim}y$$

$$f(x) = \int_y f(x,y)\,dy$$

$$f(y) = \int_x f(x,y)\,dx$$

Statistics

Fisher's t distribution

$$f(t) = \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right) \cdot \left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$\beta(m,n) = \frac{\sqrt{m} \cdot \sqrt{n}}{\sqrt{m+n}} . \qquad \sqrt{m} = (m-1)!$$

moments / constants of t disⁿ

As t disb^n symmetrical at about origin, therefore all odd order moments are zero.

Then, where, $r = 0,1,2 \ldots$

$$\mu_{2r+1} = \mu'_{2r+1} = 0$$

$\therefore \mu_1' = \text{mean} = 0.$

$$\mu_3 = \mu_5 = \mu_{2r+1} = 0$$

Now, the even ordered moments

$$\mu_{2r} = \mu'_{2r} = E(t^{2r}) = \int_t t^{2r} f(t)\, dt$$

$$\boxed{E(x^n) = \int_x x^n f(x)\, dx} \quad = \int_{-\infty}^{\infty} t^{2r} \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right) \left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\, dt$$

$$= \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{-\infty}^{\infty} \frac{t^{2r}}{\left\{1+\frac{t^2}{n}\right\}^{\frac{n+1}{2}}}\, dt$$

$$= \frac{2}{\sqrt{n}\,\beta\left(\frac{1}{2},\frac{n}{2}\right)} \cdot \int_0^{\infty} t^{2r} \frac{1}{\left(1+\frac{t^2}{2}\right)^{\frac{n+1}{2}}}\, dt$$

put. $\left(1+\frac{t^2}{n}\right) = x^{-1}$

$\Rightarrow \cancel{t^2 = (x-1)n} \; \frac{n(1-x)}{x}$

$\Rightarrow t^2 = n(1-x)/x$

$\Rightarrow 2t\,dt = -\frac{n}{x^2}$

$\therefore dt = -\frac{n}{2tx^2}\, dx$

if, $t=0$, then, $x=1$

$t=\infty$, then, $x=0$

$$= \frac{2}{\sqrt{n}\,\beta\left(\frac{1}{2},\frac{n}{2}\right)} \int_1^{0} \frac{t^{2r}}{\left(\frac{1}{x}\right)^{\frac{n+1}{2}}} \cdot \frac{-n}{2tx^2}\, dx$$

$$= \frac{\sqrt{n}}{\beta\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 \frac{t^{2r}}{\left(\frac{1}{x}\right)^{\frac{n+1}{2}} tx^2}\, dx$$

$$= \frac{\sqrt{n}}{\beta\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 (t^2)^{(2r-1)/2} \cdot x^{\{(n+1)/2\}-2}\, dx$$

$$= \frac{\sqrt{n}^{\,r}}{\beta\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 \left[\frac{n(1-x)}{x}\right]^{r-\frac{1}{2}} x^{\frac{n+1}{2}-2}\, dx$$

$$= \frac{n^r}{\beta\left(\frac{1}{2},\frac{n}{2}\right)} \int_0^1 x^{\left(\frac{n}{2}-r\right)-1} (1-x)^{\left(r+\frac{1}{2}\right)-1}$$

$$\because \boxed{\int_0^1 y^{m-1}(1-y)^{n-1}\, dx = \beta(m,n)}$$

$$\boxed{\frac{1}{\beta(m,n)}\int_0^1 y^{m-1}(1-y)^{n-1}\,dy = \beta(m,n)}$$

$$= \frac{\sqrt{n}^{\,r}}{\beta\left(\frac{1}{2},\frac{n}{2}\right)}\,\beta\left(\frac{n}{2}-r,\ r+\frac{1}{2}\right)$$

$$= n^{r}\frac{\Gamma\left(\frac{n}{2}-r\right)\Gamma\left(r+\frac{1}{2}\right)}{\cancel{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}}\times\frac{\cancel{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

$$= n^{r}\frac{\Gamma\left(\frac{n}{2}-r\right)\Gamma\left(r+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}.$$

$$= \frac{n^{r}\left(r-\frac{1}{2}\right)\left(r-\frac{3}{2}\right)\cdots\frac{3}{2}\cdot\frac{1}{2}\,\cancel{\Gamma\frac{1}{2}}\;\cancel{\Gamma\left(\frac{n}{2}-r\right)}}{\cancel{\Gamma\frac{1}{2}}\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)\cdots\left(\frac{n}{2}-r\right)\cancel{\Gamma\left(\frac{n}{2}-r\right)}}$$

$$\mu_{2r} = \frac{n^{r}(2r-1)(2r-3)\cdots 3\cdot 1}{(n-2)(n-4)\cdots\cdots(n-2r)}, \quad \boxed{\frac{n}{2}>r}$$

in particular, $r=1,2,3\ldots$

$$\mu_2 = n\left(\frac{1}{n-2}\right) = \frac{n}{n-2}$$

$$\mu_4 = n^2\frac{3.1}{(n-2)(n-4)} = \frac{3n^2}{(n-2)(n-4)} \quad n>4$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0}{\mu_2^3} = 0.$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3\left(\frac{n-2}{n-4}\right)$$

$$\lim_{n\to\infty}\beta_2 = 3\lim_{n\to\infty}\left\{\frac{1-\frac{2}{n}}{1-\frac{4}{n}}\right\} = 3.$$

For large n $(n \to \infty)$

t dis^n tends to standard normal dist^n.

proof: For n d.f. the pdf of t distn

$$f(t) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)} \left\{1 + \frac{t^2}{n}\right\}^{-\frac{n}{2}-\frac{1}{2}}$$

$$\lim_{n\to\infty} f(t) = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \cdot \frac{\Gamma\frac{n+1}{2}}{\Gamma\frac{n}{2}\Gamma\frac{1}{2}} \cdot \lim_{n\to\infty} \left[\left\{1 + \frac{t^2}{n}\right\}^n\right]^{-\frac{1}{2}}$$

$$\times \lim_{n\to\infty} \left\{1 + \frac{t^2}{2}\right\}^{-\frac{1}{2}}.$$

$$= \frac{1}{\sqrt{2\pi}} 1 \cdot e^{-\frac{t^2}{2}} \quad \lim_{n\to\infty} f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad -t<t<\infty$$

$X \sim N(0,1)$

$f(x) \sim t$.

F-statistics

$$\boxed{F = \frac{X/n_1}{Y/n_2}}$$

$X \sim \chi^2$ — dis^n with n_1 df

and $Y \sim \chi^2$ " " n_2 df

F dis^n with (n_1, n_2) d.f.

If X and Y are statistically independent then

which has the following pdf as $f(F) = \dfrac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} F^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1 + \frac{n_1 F}{n_2}\right)^{\frac{n_1+n_2}{2}}}$

$F \geq 0$.

CS CamScanner

Application

(1) Testing for equality of two population variances
(2) " " significance of an observed multiple correlation coefficient.

(3) " " " observed sample correlation

(4) " " the linearity of Regression

5) " " for equality several mean.
6) " " " ANOVA (Analysis of variance)

11 $\boxed{2}$ 3

~~10~~9. 20

$\alpha/2$

20

5 5 6

12 12 $\boxed{2}$ 3

11×2=22

2×10=~~18~~

$\boxed{10}$

$\boxed{10}$

1 5 1 2

a b

10

(n-1)

4

1 → 1

~~2~~ 2

statistics :

$$f(F) = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \cdot \frac{(F)^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} \; ; \; F \geq 0$$

$$\int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}} dx = \beta(m,n)$$

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

Mean(μ): $\mu_1' = E(F)$

Variance: $\mu_2 = V(F) = \mu_2' - (\mu_1')^2$

$$= E(F^2) - \{E(F)\}^2$$

$$\boxed{E(x) = \int_x x \, f(x) \, dx}$$

$$E(F) = \int_F F f(F) \, d\cancel{x}F$$

$$= \cancel{\int_F} \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{F \, (F)^{\left(\frac{n_1}{2}+1\right)-1}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} \cdot dF$$

$\Rightarrow$ put $\frac{n_1}{n_2}F = x$

$\Rightarrow F = \frac{n_2}{n_1}x$

$\therefore dF = \frac{n_2}{n_1}dx$

$$= \int_0^\infty \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \cdot \int_0^\infty \frac{\left\{\left(\frac{n_2}{n_1}\right)x\right\}^{\left(\frac{n_1}{2}+1\right)-1}}{(1+x)^{\frac{n_1+n_2}{2}}} \cdot \frac{n_2}{n_1}dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \cdot \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}+1-1+1} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+1\right)-1}}{(1+x)^{\frac{n_1}{2}+\frac{n_2}{2}}}dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}+1} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+1\right)-1}}{(1+x)^{\left(\frac{n_1}{2}+1\right)+\left(\frac{n_2}{2}-1\right)}} \cdot dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}+1} \beta\left(\frac{n_1}{2}+1, \frac{n_2}{2}-1\right)$$

$$= \frac{n_2}{n_1} \frac{\cancel{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}}{\Gamma\frac{n_1}{2}\,\Gamma\frac{n_2}{2}} \cdot \frac{\Gamma\left(\frac{n_1}{2}+1\right)\Gamma\left(\frac{n_2}{2}-1\right)}{\cancel{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}}$$

$$= \frac{n_2}{\cancel{n_1}} \frac{\frac{\cancel{n_1}}{2}\cancel{\Gamma\frac{n_1}{2}} \cdot \cancel{\Gamma\left(\frac{n_2}{2}-1\right)}}{\cancel{\Gamma\frac{n_1}{2}} \cdot \left(\frac{n_2}{2}-1\right)\cancel{\Gamma\left(\frac{n_2}{2}-1\right)}}$$

$$= \frac{n_2}{2} \cdot \frac{1}{\frac{n_2-2}{2}} = \frac{n_2}{\cancel{2}} \times \frac{\cancel{2}}{n_2-2} = \frac{n_2}{n_2-2}, \quad n_2 > 2$$

$$\left(\frac{n_1}{2}+2\right) \quad \frac{n_2}{2}-2 \quad \frac{n_1+n}{2} \qquad \frac{n_1}{2}+2$$

$$E(F) = \int_F f(F)\,dF$$

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

$$\mu_2' = E(F^2) = \int_0^\infty F^2 f(F)\cdot dF$$

put $\frac{n_1}{n_2} F = x$

$F = \frac{n_2}{n_1} x$

$\therefore dF = \frac{n_2}{n_1} dx.$

$\Gamma(n+2) = n(n+1)\Gamma n$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{F^2 (F)^{\left\{\frac{n_1}{2}+1\right\}-1}}{\left(1+\frac{n_1}{n_2} F\right)^{\frac{n_1+n_2}{2}}} \cdot dF.$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{\left(\frac{n_2}{n_1}x\right)^{\left(\frac{n_1}{2}+2\right)-1}}{(1+x)^{\frac{n_1+n_2}{2}}} dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}+2}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+2\right)-1}}{(1+x)^{\frac{n_1}{2}+\frac{n_2}{2}}} dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^2}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+2\right)-1}}{(1+x)^{\left(\frac{n_1}{2}+2\right)+\left(\frac{n_2}{2}-2\right)}} dx$$

$$\boxed{= \frac{\left(\frac{n_2}{n_1}\right)^2}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \beta\left(\frac{n_1}{2}+2, \frac{n_2}{2}-2\right)}$$

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

$$= \frac{\left(\frac{n_2}{n_1}\right)^2}{\frac{\Gamma\frac{n_1}{2}\Gamma\frac{n_2}{2}}{\Gamma\frac{n_1}{2}+\frac{n_2}{2}}} \cdot \frac{\Gamma\frac{n_1}{2}+2\;\Gamma\frac{n_2}{2}-2}{\Gamma\frac{n_1}{2}+\frac{n_2}{2}}$$

$$\left(\frac{n_1}{2}+1\right)\left(\frac{n_1}{2}\right)\Gamma\frac{n_1}{2}$$

$$\left(\frac{n_2}{2}-1\right)\left(\frac{n_2}{2}-2\right)\Gamma\frac{n_2}{2}-2$$

$$\frac{\left(\frac{n_2}{n_1}\right)^2}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\,\beta\left(\frac{n_1}{2}+2,\ \frac{n_2}{2}-2\right)$$

$$=\frac{\left(\frac{n_2}{n_1}\right)^2\cancel{\Gamma\frac{n_1+n_2}{2}}}{\Gamma\frac{n_1}{2}\ \Gamma\frac{n_2}{2}}\times\frac{\Gamma\left(\frac{n_1}{2}+2\right)\cdot\Gamma\left(\frac{n_2}{2}-2\right)}{\cancel{\Gamma\frac{n_1+n_2}{2}}}$$

$$=\frac{\left(\frac{n_2}{n_1}\right)^2}{\cancel{\Gamma\frac{n_1}{2}}\ \Gamma\frac{n_2}{2}}\cdot\left(\frac{n_1}{2}+1\right)\frac{n_1}{2}\cdot\cancel{\Gamma\frac{n_1}{2}}\ \Gamma\left(\frac{n_2}{2}-2\right)$$

$$=\frac{\left(\frac{n_2}{n_1}\right)^2\left(\frac{n_1}{2}+1\right)\frac{n_1}{2}\cdot\cancel{\Gamma\left(\frac{n_2}{2}-2\right)}}{\left(\frac{n_2}{2}-1\right)\left(\frac{n_2}{2}-2\right)\cancel{\Gamma\left(\frac{n_2}{2}-2\right)}}$$

$$=\frac{\frac{n_1}{2}\left(\frac{n_2}{n_1}\right)^2\left(\frac{n_1}{2}+1\right)}{\left(\frac{n_2}{2}-1\right)\left(\frac{n_2}{2}-2\right)}=\frac{\frac{n_1}{2}\cdot\frac{n_2^2}{n_1^2}\cdot\left(\frac{n_1}{2}+1\right)}{\frac{n_2^2}{4}-n_2-\frac{n_2}{2}+2}$$

$$=\frac{\frac{n_2^2}{4}+\frac{n_2^2}{2n_1}}{\frac{n_2^2-4n_2-2n_2+8}{4}}=\frac{\frac{n_1n_2^2+2n_2^2}{n_1}}{n_2(n_2-4)-2(n_2-4)}$$

$$=\frac{n_2^2+2\frac{n_2^2}{n_1}}{n_2^2-4n_2-2n_2+8}=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}\qquad n>$$

04/02/2024

Statistics

Estimation

point estimation — single value

interval estimation — range

(i) MLE

(ii)

x — Random

x_1 — p_1

x_2 — p_2

-- -- --

x_n - - - p_n

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

$$E(x) = \mu = \lambda$$

estimated value

$(\bar{x} - \mu) \neq 0$ — reality

$\Rightarrow \boxed{\bar{x}} = \hat{\mu}$

estimator

$\bar{x} = \sum x_i / n$

statistics

*what is estimator?

*** ~~Good~~ Best estimator four condition

(i) Consistency — $n \to \infty$ হলে $\bar{x} \to \mu$

(ii) Unbiasedness — If $E(\bar{x}) = \mu$ then $\bar{x}$ is unbiased

(iii) Efficiency

(iv) Sufficiency

Consistency

An estimator $t_n = (x_1, x_2, \ldots \ldots x_n)$ based on a random sample of size n is said to be consistent estimator of the parameter

$g(\theta)$, $\theta \in \Omega$ (Ω parameter space)

if t_n $\boxed{\text{converges to}}$ $g(\theta)$. in probability such that
tends to

if $t_n \xrightarrow{p} g(\theta)$.

In other words. t_n is consistent estimator of $g(\theta)$ if ever $\epsilon > 0$ & $\eta > 0$ there exists a positive integer. $n \geq m$ (where m is large) such that $p[\{t_n - g(\theta)\} < \epsilon] \to 1$ as $n \to \infty$

$\Rightarrow p[\{t_n - g(\theta)\}] < \epsilon] > 1 - \eta \quad \forall\ n \geq$

In ~~pr~~ Ex. if $x_1, x_2, \ldots \ldots x_n$ be a r.s of size n from a population with finite mean, $E(x_i) = \mu$, then the weak law of large number, we have

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \xrightarrow{p} E(x_i) = \mu \text{ as } n \to \infty.$$

Hence
Definition
statistics, Parameter, random variable, estimator, estimation

Unbiased estimator

An estimator of a given parameter (θ) is said to be an unbiased if its expected value is equal to the true value of the parameter.

In other words, an estimator is unbiased if it produces parameter estimates that are on average correct.

If t_n be a statistic, calculated from a sample $x_1, x_2 \dots x_n$ of size n from a density $f(x, \theta)$ such that $E(t_n) = \theta$ then t_n is an unbiased estimator of θ.

Example

If (x follows) $x \sim N(\mu, \sigma^2)$ variate then sampling mean $\bar{x}$ is an unbiased estimator of μ is

$$E(\bar{x}) = \mu$$

we have to prove $E(\bar{x}) = \mu$.

we know, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$E(\bar{x}) = E\left(\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\}\right)$$

$$= \frac{1}{n} E\left(\left\{\sum_{i=1}^{n} x_i\right\}\right)$$

$$\Rightarrow E(\bar{x}) = E\left\{\frac{1}{n}\right\} E\left(x_1 + x_2 + x_3 + \dots + x_n\right)\}$$

$$\Rightarrow E(\bar{x}) = \frac{1}{n}\left[E(x_1) + E(x_2) + \dots + E(x_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \dots + \mu\right] \quad (\Sigma x_i = \mu)$$

$$= \frac{1}{n} \cdot n\mu$$

$$\therefore E(\bar{x}) = \mu.$$

~~##~~ When the roots of auxiliary equation has some repeated roots.

$$(D-m_1)^2 y = 0$$

Let, $(D-m_1)y = u$

$\Rightarrow (D-m_1) = u$

$\Rightarrow Du - m_1 u = 0$

$\Rightarrow \frac{du}{dx} - m_1 u = 0$

6/04/2024

statistics

Ex: If t is a unbiased estimator of θ, then t^2 is a unbiased estimator for θ^2.

Proof:

As t is an unbiased estimator of θ then we get,

$E(t) = \theta$ ------ (i)

We know from definition of variance,

$$V(t) = E(t^2) - \{E(t)\}^2$$
$$= E(t^2) - \{\theta\}^2$$
$$= E(t^2) - \theta^2$$

$E(t^2) = V(t) + \theta^2$

$\therefore E(t^2) \neq \theta^2$ as $V(t) > 0$

$\therefore t^2$ is not unbiased for θ^2

Thus t^2 is a biased estimator.//.

Ex-2

A random sample $(x_1, x_2, x_3, x_4, x_5)$ of size 5 is from a normal popn with unknown mean μ. Consider the following estimators to estimate μ.

(i) $t_1 = \frac{x_1 + \cdots + x_5}{5}$,

(ii) $t_2 = \frac{x_1 + x_2}{2} + x_3$

(iii) Find λ for unbiasedness of $t_3 = \frac{2x_1 + x_2 + \lambda x_3}{3}$

(iv) t_1, t_2 is unbiased or not?

(i) $E(x_i) = \mu$, $i = 1,2,3,4,5$

$$E(t_1) = E\left(\frac{x_1+x_2+x_3+x_4+x_5}{5}\right)$$
$$= \frac{1}{5} E(x_1+x_2+x_3+x_4+x_5)$$
$$= \frac{1}{5} \cdot 5\mu$$
$$= \mu.$$

$\therefore E(x_i) = E(t_1) = \mu$.

So, t_1 is an unbiased estimator for μ.

(ii) $t_2 = \frac{x_1+x_2}{2} + x_3$

$$E(t_2) = E\left(\frac{x_1+x_2}{2} + x_3\right)$$

$$= \frac{E\left(x_1\right)+E\left(x_2\right)}{2} + E(x_3)$$

$$= \frac{\mu+\mu}{2} + \mu$$

$$= 2\mu$$

As $E(t_2) \neq \mu$

$\therefore$ t_2 is a biased estimator for μ.

(iii) $t_3 = \frac{2x_1+x_2+\lambda x_3}{3}$

$$\Rightarrow E(t_3) = \frac{1}{3}\{2E(x_1)+E(x_2)+\lambda E(x_3)\}$$

$$= \frac{1}{3}(2\mu+\mu+\lambda\mu)$$

As t_3 is an unbiased estimator,

$\therefore$ $E(t_3) = \mu$

$$\Rightarrow \frac{1}{3}(2\mu+\mu+\lambda\mu) = \mu$$

$$\Rightarrow 3\mu+\lambda\mu = 3\mu$$

$$\Rightarrow \lambda\mu = 0$$

$\therefore$ $\lambda = 0$ as $\mu \neq 0$

(Ans).

1. Definition of point estimation

(ii) method of "

(iii) principles of ML

(iv) Applications of ML

Efficiency:

For two unbiased estimator t_1 and t_2 for θ, then their variances are $V(t_1)$ and $V(t_2)$.

If $V(t_1) < V(t_2)$ then t_1 is more efficient.

* $V(t_1) \doteq V(t_2)$ none is efficient.

* suppose for t_i $\{i=1,2,\ldots,k\}$ all are unbiased estimator for a given parameter $g(\theta)$ and their respective variances $V(t_i)$.

Now, $E = \dfrac{V(t_i)}{V(t_j)}$ $\{j \neq i\}$

if $E<1$, then t_i is the most efficient estimator for $g(\theta)$.

Minimum Variance Bound

$E = \dfrac{MVB}{V(t_i)}$; $i=1,2,\ldots,k$

if $E<1$, then t_i is the most efficient estimator $g(\theta)$.

for each t_i, maximum value of E in this case is most efficient

Sufficiency

A good estimator has 4 criteria:

1) Unbiasedness
2) Consistency
3) Efficiency
4) Sufficiency

Efficiency

$V(t_1) < V(t_2)$ where t_1 and t_2 are two unbiased estimators of θ, then t_1 is more efficient.

MVB — Minimum Variance Bound

$$E = \frac{MVB}{V(t_i)}\ ;\ i = 1, 2, \dots, K$$

if E is maximum (closest to 1), efficiency is maximum.

** Cramer–Raw Inequality

Statement:

If $T = t(x_1, x_2, \dots x_n)$ be an unbiased estimator for $g(\theta)$ from pdf $f(x, \theta)$, then C-R inequality

$$V(t) \geqslant \frac{\left\{\frac{d}{d\theta}\, g(\theta)\right\}^2}{E\left\{\frac{\delta}{\delta\theta} \log L\right\}^2} = \frac{\{g'(\theta)\}^2}{I(\theta)}$$

Example:

$f(x_1, \theta) = \theta x_1^{\theta} \qquad 0 < x < 1$

$f(x_2, \theta) = \theta x_2^{\theta}$

$L(x, \theta) = f(x_1, \theta)\, f(x_2, \theta) \dots f(x_n, \theta)$

প্রতিটি pdf এর গুণফল, আসলেই Likelyhood function.

$L \rightarrow$ Likelyhood function of the parameter.

Joint density Function is the function of sample observation

Sufficiency

An estimator is said to be sufficient for a parameter if it contains all the information in the sample regarding the parameters.

If $T = t(x_1, x_2, \dots, x_n)$ is an unbiased estimator of a parameter θ, based on a sample $x_1, x_2, \dots, x_n$ of size n from the population with pdf $f(x, \theta)$ such that the conditional distribution of $x_1, x_2, \dots, x_n$ given T is independent of θ, then T is a sufficient estimator for θ.

Factorization Theorem or Neyman Theorem

Statement:

If $T = t(x_1, x_2, \ldots, x_n)$ is sufficient estimator for $g(\theta)$ if and only if the joint density function L (say) of the sample values can be expressed as in the form

$L = g_\theta[t(x)]\, h(x)$

where $g_\theta[t(x)]$ depends on θ ~~but~~ and x but $h(x)$ is independent of θ.

Example:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from $N(\mu, \sigma^2)$. Find sufficient estimators for μ and σ^2.

Solution:

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

$$X \sim N(0,\sigma^2) = \frac{e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}}{\sigma\sqrt{2\pi}} = \frac{e^{-\frac{1}{2\sigma^2}x^2}}{\sigma\sqrt{2\pi}}$$

$$X \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\therefore L = f(x_1;\mu,\sigma^2)\ldots f(x_n;\mu,\sigma^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}$$

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

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

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

$$= g_\theta[t(x)] \cdot h(x)$$

$$= \cancel{g(\theta)}$$

where,

$$g_\theta[t(x)] = (2\pi\sigma^2)^{-\frac{n}{2}} e^{\left[-\frac{1}{2\sigma^2}\right\} t_2(x) - 2\mu t_1(x) + n\mu^2}$$

$$t(x) = \{t_1(x), t_2(x)\} = \left\{\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right\} \text{ and } h(x)$$

Therefore, by the factorization theorem $t_1(x) = \sum_{i=1}^{n} x_i$ is the sufficient statistic for μ.

and $t_2(x) = \sum_{i=1}^{n} x_i^2$ is the sufficient statistic for σ^2

Ex: $f(x,\theta) = \theta x^{\theta-1}$ $0<x<1$

Show that $t_1 = \prod_{i=1}^{n} x_i$ is the sufficient estimator for θ.

Solution:

$L = f(x_1,\theta)\, f(x_2,\theta) \dots f(x_n,\theta)$

$= \theta x_1^{\theta-1} \cdot \theta x_2^{\theta-1} \dots \theta x_n^{\theta-1}$

$= \theta^n (x_1 \cdot x_2 \cdot x_3 \dots x_n)^{\theta-1}$

$= \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}$

$= \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta} \cdot \frac{1}{\prod_{i=1}^{n} x_i}$

$= g_\theta(t(x)) \cdot h(x)$

where,

$g_\theta(t(x)) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}$

is the function of θ and x but $h(x)$ is independent of θ. Therefore, by factorization theorem.

$t(x) = \prod_{i=1}^{n} x_i$ is the sufficient estimator for θ.

Q

1. Define estimation, point estimation.
2. Write down the methods of point estimation.
3. Explain the principle of MLE (Maximum Likelihood Estimation)
4. Properties of MLE
5. Mathematical example.

Definition of Point estimation

In statistics point estimation is a process of finding an approximate value of some parameters such as mean (μ) of a population from the random sample of the population.

Methods of point estimation

(I) Method of maximum likelihood estimation (MLE)
(II) Method of minimum variance
(III) method of moments.
(IV) method of least squares.
(V) Method of minimum χ^2
(VI) method of inverse probability

Principals of MLE

~~suppose~~ let $x_1\ x_2\ x_n$ be a random sample of size n drawn from a given population with pdf $f(x;\theta)$ where θ is the population parameter (s).

Now the likelihood function denoted by $L(x/\theta) = L($

$= f(x_1;\theta) f(x_2,\theta) \cdots f(x_n,\theta)$

$$L(x/\theta) = \prod_{i=1}^{n} f(x_i;\theta) \quad \text{---} \quad (1)$$

To maximize (1), we have to get $\frac{\delta L(x;\theta)}{\delta\theta} = 0$ —— (2)

$$\Rightarrow \frac{\delta}{\delta\theta}\left\{\prod_{i=1}^{n} f(x_i;\theta)\right\} = 0$$

Again, $$\frac{\delta^2 L(x;\theta)}{\delta\theta^2} = \frac{\delta^2}{\delta\theta^2}\left\{\prod_{i=1}^{n} f(x_i;\theta)\right\}$$

If $\frac{\delta^2 L(x;\theta)}{\delta\theta^2} < 0$, then the value of θ obtained from (2) is the maximum value of θ. which is required MLE for θ.

For simplicity due to exponentiation form of likelihood function we may take log. of $L(x;\theta)$.

$$\frac{\delta \, Log(x;\theta)}{\delta\theta} = 0 \text{ —— } (1)$$

$$\frac{\delta^2}{\delta\theta^2} \log L(x;\theta) < 0.$$

Example:

1. $f(x;\lambda) = \lambda e^{-\lambda x}$; $x \geq 0$
2. $f(x;\theta) = \frac{1}{\theta} \cdot e^{-x/\theta}$; $x \geq 0$
3. $f(x,m) = \frac{e^{-m} m^x}{x!}$; $x = 0, 1, 2, 3, \ldots, n$
4. $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2}$

$-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma^2 \geq 0$

Sol

Given,

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

Likelihood function

$$L(x;\lambda) = f(x_1,\lambda)\cdots f(x_n,\lambda).$$

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

$$= e^{-n\lambda}\left\{\frac{\lambda^{x_1+x_2+\cdots x_n}}{\prod_{i=1}^{n}(x_i)!}\right\}$$

Let $\hat{\lambda}$ be the estimated value of λ.

Taking log on both sides,

$$\log L(x;\lambda) = -n\hat{\lambda} + \sum x_i \log\hat{\lambda} - \sum_{i=1}^{n}\log$$

$$\frac{\delta}{\delta\lambda}\log L = 0$$

$$\Rightarrow -n + \frac{\sum x_i}{\hat{\lambda}} = 0$$

$$\Rightarrow \frac{\sum x_i}{\hat{\lambda}} = n$$

$$\Rightarrow \sum x_i = n\hat{\lambda}.$$

$$\therefore \hat{\lambda} = \frac{\sum x_i}{n} = \bar{x}$$

5

Given,

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

Likelihood function

$$L(x;\lambda) = f(x_1,\lambda) \cdots f(x_n,\lambda)$$

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

$$= e^{-n\lambda}\left\{\frac{\lambda^{x_1+x_2+\cdots x_n}}{\prod_{i=1}^{n}(x_i)!}\right\}$$

Let $\hat{\lambda}$ be the estimated value of λ.

Taking log on both sides,

$$\log L(x;\lambda) = -n\hat{\lambda} + \sum x_i \log\hat{\lambda} - \sum_{i=1}^{n}\log(x_i!)$$

$$\frac{\delta}{\delta\lambda}\log L = 0$$

$$\Rightarrow -n + \frac{\sum x_i}{\hat{\lambda}} = 0$$

$$\Rightarrow \frac{\sum x_i}{\hat{\lambda}} = n$$

$$\Rightarrow \sum x_i = n\hat{\lambda}.$$

$$\therefore \hat{\lambda} = \frac{\sum x_i}{n} = \bar{x}$$

$$\frac{\delta \log L}{\delta \lambda} = n\left(\frac{\bar{x}}{\lambda} - 1\right) = g(x, \lambda) \cdot h(x)$$

Here; $g(x,\lambda)$ depends on x and λ but $h(x)$ is independenof λ.

Therefore by factorization theorem, $\bar{x}$ is a sufficient estimator for λ.

~~* If (x follows) $x \sim N(\mu, \sigma^2)$ variate then sample mean $\bar{x}$ is an unbiased estimator for λ is $E(\bar{x}) = \lambda$.~~

~~we have to prove. $E(\bar{x}) = \lambda$.~~

~~we know, $\bar{x} = \frac{1}{n}$.~~

* Prove $E(\bar{x}) = \lambda$ is an unbiased estimation.

We have to proof that.

$$E(\bar{x}) = \lambda \ldots$$

$$E(\bar{x}) = E\left(\frac{\sum x_i}{n}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(x_i) \quad \text{——— (2)}$$

$$E(x_i) = \sum_{i=1}^{\infty} x \cdot p(x) = \sum_{i=1}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$= \sum_{i=1}^{\infty} \cancel{x_i} \frac{e^{-\lambda} \lambda^{x_i - 1} \cdot \lambda}{\cancel{x_i}(x_i - 1)!}$$

$$= \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda} \lambda^{x_i - 1}}{(x_i - 1)!} = \lambda \cdot 1 = \lambda.$$

<u>Statistics</u>

<u>Section-A</u>

	Question Pattern
1. Sampling Distribution (χ^2, t, F)	(1/2)
2. Estimation criteria	(1/0)
3. Point estimation	(1/0)

<u>Section-B</u>

1. Contingency Table	1
2. Test of Hypotheses	1
3. Non parametric test	1

Hypotheses Testing — Section B
+
Estimation

= Inferences

*Relationship between two Parents and children heights using 2x2 contingency tables : for 50 test cases.

Parents

		Yes	No
Child	Yes	30	20
	No	20	30

H_0 : There is no relationship between parents and childs height.

H_1 :

If cal (χ^2-test or t-test or F-test) $\geq$ Given (χ^2/t/F test) then we reject H_0.

12/01/24

* nodeptr;

Statistics

$\lim_{n \to \infty} \chi^2 \longrightarrow$ Normal dis'n

$(n \geq 30)$

MGF $n \to \infty \longrightarrow$ "

if $x \sim \chi^2_{(n)}$

$$M_x(t) = (1-2t)^{-\frac{n}{2}}$$

The standard MGF of χ^2-variate z is $M_{\frac{x-\mu}{\sigma}}(t)$

$= M_{\left(\frac{x}{\sigma}-\frac{\mu}{\sigma}\right)}(t)$

$= M_{\frac{x}{\sigma}}(t) + M_{\frac{\mu}{\sigma}}(t$

* $M_{(x_1 \pm x_2)}(t) = M_{x_1}(t)\, M_{x_2}(t)$