Statistical Inference

Population → unknown parameters ⇒ we try to estimate.

sample → sample statistics

point estimation

interval estimation

Testing of hypothesis

population → θ unknown

t ← sample statistics

point estimation

$$f(x;\theta)$$

sample

trying to estimate.

$x_1, x_2 \dots x_n$ sample ⇒ t

t is an estimator of θ

when can you say t is satisfactory

⇒ if $|t-\theta|$ is as small as possible

Properties of a good estimator.

1. Unbiasedness $E(t) = \theta$

(1) Unbiasedness $E(t) = \theta$

or $E(t) - \theta = 0$

if $E(t) \neq \theta$

$E(t) - \theta \neq 0$

(then t is biased) ✗

for example

Normal population

θ is μ population mean

t is $\bar{x}$ → sample mean

You can say that $\bar{x}$ is an unbiased estimator of μ if

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

(2) Min variance

two sample sets

t_1 t_2 θ

if t_1 and t_2 both are unbiased then we will check $var(t_1)$ and $var(t_2)$

if $v(t_1) < v(t_2)$

then t_1 is a better estimator of θ.

(*) we call it a Min variance Unbiased Estimator (MVUE)

(3) Consistency: $P\{|t_n - \theta| > \epsilon\} \Rightarrow 0$

(3) Consistency: $P\{|t_n - \theta| > \epsilon\} \Rightarrow 0$

t_1 and t_2

as $n \to \infty$

value of t_n converges to θ as $n \to \infty$.

(4) Efficiency: (i) $E(t_n) \to \theta$.

(ii) $var(t_n) \to 0$

as $n \to \infty$

then the statistic t is called the efficient estimator of θ if t is asymptotically normally distributed and $avar(t_1) \le avar(t_2)$

this is ~~sufficient~~ efficient estimator.

(v) Sufficiency: $X_1, X_2 \cdots X_N$ with unknown θ.

$x_1, x_2 \cdots x_n$ sample size n

then if you can write,

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

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

Numerical

if the variance of independent and unbiased estimators T_1, T_2, T_3 of θ are in the ratio 2:3:5, which of the following estimators of θ would you prefer most?

$t_1 = (2T_1 + T_2 + T_3)/4$, $t_2 = (T_1 + T_2 + 2T_3)/4$
$t_3 = (T_1 + 2T_2 + T_3)/4$

$T_1 \quad T_2 \quad T_3$ of θ

$$E(T_1) = E(T_2) = E(T_3) = \theta$$

$$\frac{V(T_1)}{2} = \frac{Var(T_2)}{3} = \frac{Var(T_3)}{5} = k$$

$$E\left(\frac{2T_1 + T_2 + T_3}{4}\right) = \frac{2E(T_1)}{4} + \frac{E(T_2)}{4} + \frac{E(T_3)}{4}$$

$$= \cancel{2}\,\theta + \theta + \theta$$

$$= \frac{\theta}{2} + \frac{\theta}{4} + \frac{\theta}{4}$$

$$= \theta/2 + 2\theta/4 = \frac{\theta}{2} + \frac{\theta}{2} = \theta$$

$$E\left(\frac{T_1+T_2+2T_3}{4}\right) = \frac{\theta}{4} + \frac{\theta}{4} + \frac{2\theta}{4} = \theta$$

$$E(T_1+2T_2+T_3)/4 = \frac{\theta}{4} + \frac{2}{4}\theta + \frac{\theta}{4} = \theta$$

Thus all three estimators are unbiased for θ.

$$V\left(\frac{2T_1+T_2+T_3}{4}\right) = \frac{1}{16}\left[4V(T_1) + V(T_2) + V(T_3)\right]$$

$$= \frac{1}{16}\left[4\times 2k + 3k + 5k\right]$$

$$= k$$

$$V(T_1+T_2+2T_3/4) = \frac{2k+3k+4\times 5k}{16}$$

$$= \frac{25k}{16} > k$$

$$V\left(\frac{T_1+2T_2+T_3}{4}\right) = \frac{2k+4\times 3k+5k}{16}$$

$$= \frac{19}{16} k$$

$$k < \frac{19}{16} k < \frac{25}{16} k$$

$V(2T_1 + T_2 + T_3)/4$ is min

$\therefore$ the estimator $\frac{2T_1 + T_2 + T_3}{4}$ is the preferred statistics among the given three.

Maximum Likelihood Method / estimation (MLE).

$\downarrow$

maximisation of a log likelihood fn.

$X_1, X_2 \ldots X_N \to \theta$

$x_1, x_2 \ldots x_n \to t$

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

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

f.o.c $\left.\frac{\partial L(\theta)}{\partial \theta}\right|_{\theta = \hat{\theta}} = 0$

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0 \ldots$$

$$\left.\frac{\partial^2 L(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0 \quad \text{for maximisation}$$

Ex Let us consider a set of n Bernoullian trials with p as probability of success in a trial.

With the ith trial we associate a variable x_i such that,

$$x_i = \begin{cases} 1 & \text{if there is success} \\ 0 & \text{otherwise.} \end{cases}$$

$p^{x_1} p^{x_2} \ldots p^{x_n}$

$p^{\Sigma x_i}$

Then the pmf of x_i is

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

$$L(p) = \prod_{i=1}^{n} f(x_i; p) = p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{\Sigma(1-x_i)}$$

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

$$\log_e L(p) = \left(\Sigma x_i\right) \log_e p - (n - \Sigma x_i) \log_e (1-p)$$

F.O.C $$\frac{d \log_e L(p)}{dp} = \frac{\Sigma x_i}{p} - \frac{n - \Sigma x_i}{1-p}$$

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

$$= \frac{\Sigma x_i - \cancel{p\Sigma x_i} - np + \cancel{p\Sigma x_i}}{p(1-p)}$$

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

F.O.C $$\frac{\partial \log L(p)}{\partial p} = 0$$

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

$\therefore \Sigma x_i - np = 0 \Rightarrow p = \frac{\Sigma x_i}{n}$

$\hat{p} = \bar{x}$

sample proportion of sum

$\therefore$ m.l.e of $\hat{p}$ is $\bar{x}$.

ans

HW: TRY to find out the mle in

HW : TRY to find out the mle in case of Normal distribution and Poison distribution

Topic : Interval estimation

$x_1, x_2 \cdots x_n$ if $\bar{x}$ is sample mean for a normal population with population mean μ and σ.

Case 1: estimate μ (unknown), σ is known.

test statistic

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Z is a standard normal deviate (distribution)

What is a standard normal distribution?

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

$\hookrightarrow$ in a linear fn

$$Z \sim \frac{X - \mu}{\sigma}$$

$$P\left[Z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right] = 1-\alpha$$

α is called the level of significance $\rightarrow$ probability of lying in rejection region

| e. std normal variable.

$Z = \frac{\quad}{\sigma}$

↳ std normal variable.

$Z \sim N(0, 1)$

$\sum Z^2 \sim \chi^2(n)$

$X \sim Z$

$Y \sim \chi^2(n)$

$= \frac{X}{Y/\sqrt{n}} \sim t_{(n)}$ statistics

$\frac{X/n}{Y/m}$ where $X \sim \chi^2(n)$

$Y \sim \chi^2(m)$

then $\frac{X/n}{Y/m} \sim F_{(m,n)}$

① μ — with σ known $\to Z$

↘ with σ unknown $\to t$

1 question

$P\left[\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \le \mu \le \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right] = 1 - \alpha$

lower limit

upper limit

Case 2: Confidence limit to μ when σ is also unknown

$\bar{x}$

$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$

or $s'^2 = \frac{1}{(n-1)} \sum (x_i - \bar{x})^2$

sample variance.

test statistics is

$$\frac{\bar{x} - \mu}{s'/\sqrt{n}} \sim t_{(n-1)}$$

$$P\left[t_{1-\frac{\alpha}{2}, n-1} \le \frac{\bar{x}-\mu}{s'/\sqrt{n}} \le t_{\frac{\alpha}{2}, n-1}\right] = 1-\alpha$$

$$P\left[\bar{x} - \frac{s'}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1} \le \mu \le \bar{x} + \frac{s'}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}\right] = 1-\alpha$$

Case 3 confidence interval of σ, when μ is known

test statistics is $\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}$

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

$$P\left[\chi^2_{1-\alpha/2, n} \le \frac{\sum (x_i - \mu)^2}{\sigma^2} \le \chi^2_{\alpha/2, n}\right] = 1-\alpha$$

$$P\left[\frac{\sum (x_i - \mu)^2}{\chi^2_{\alpha/2, n}} \le \sigma^2 \le \frac{\sum (x_i - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}, n}}\right] = 1-\alpha$$

Case 4 <u>confidence limit to σ when μ is unknown</u>

sample mean

test statistics is $\sum (x_i - \bar{x})^2$

test statistics is $\dfrac{\sum (x_i - \bar{x})^2}{\sigma^2}$

$s'^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

$$= (n-1) s'^2 / \sigma^2 \sim \chi^2_{(n-1)}$$

Confidence limits are

$$\frac{(n-1) s'^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \quad \text{and} \quad \frac{(n-1) s'^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}$$

Summary!

$\mu \rightarrow$ with σ known $\longrightarrow$ Z

$\mu \rightarrow$ σ unknown (sample) $\longrightarrow$ $t_{(n-1)}$

$\sigma \rightarrow$ with known μ $\longrightarrow$ $\chi^2_{(n)}$

$\sigma \rightarrow$ with unknown μ (sample) $\longrightarrow$ $\chi^2_{(n-1)}$

Q Suppose that a random variable of size 10, drawn from a normal population [illegible] 200 and sd 12. find

10, drawn ... has mean 40 and sd 12. find 99% confidence limits for the population mean.

(Given $t_{0.005,9} = 3.25$, $t_{0.005,10} = 2.21$)

$P(\ ) = 1-\alpha$

$\alpha = 100 - 99\%$

$= 1\%$

$\alpha = 0.01$

μ ?

confidence interval for μ

$\alpha = 0.01$

$n = 10$ $\quad \bar{x} = 40$, $s = 12$

Which case? and then calculate.