Interval estimation

Definition

An INTERVAL ESTIMATE of a real-valued parameter θ is any pair $L(x)$ and $U(x)$ where $L(x) \leq U(x)$ and x is your observed sample.

An INTERVAL ESTIMATOR is the same, but for random variables $X = (X_1, X_2, \dots, X_n)$ is the pair $L(X)$ and $U(X)$

Point estimator: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ ← random variables

Point estimate: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ ← observed values

We infer: if X is observed as x

then $L(x) \leq \theta \leq U(x)$

Definition 1.5.14

If the interval estimator $[L(X), U(X)]$

is designed so that $L(X) \leq U(X)$

and $P_\theta \left(L(X) \leq \underset{\substack{\uparrow \\ \text{unknown parameter value}}}{\theta} \leq U(X)\right) \geq 1 - \overset{\substack{\text{alpha} \\ \downarrow}}{\alpha}$

for every possible value of θ
and for some $\alpha \in (0,1)$
then $[L(X), U(X)]$ is called a

$1-\alpha$ CONFIDENCE INTERVAL

Remark: $\alpha = 0.05$
$\Rightarrow$ 0.95 confidence interval

However more common
95% confidence interval

$100(1-\alpha)$% CONFIDENCE INTERVAL

Example:

Suppose $X_1, X_2, \ldots, X_{10}$ are our random variables, observed as in the example.

Assume all $X_1, \ldots, X_{10}$ follow distribution F_X and are independent.

We are interested in unknown mean θ

Assume known that for F_X $\sigma^2 = 27.04$

<u>Chebyshev's inequality</u>

$$P(|Y - E[Y]| < k\sqrt{Var(Y)}) \geq 1 - 1/k^2$$

for all $k > 0$

$$\Rightarrow P(Y - k\sqrt{Var(Y)} < E[Y] < Y + k\sqrt{Var(Y)}) \geq 1 - 1/k^2$$

$$Y = \bar{X} \Rightarrow E[X] = \theta$$

$$Var(\bar{X}) = \frac{\sigma^2}{n} \quad \text{(Prop 1.2.6)}$$

$$P\left(\bar{X} - k\frac{\sigma}{\sqrt{n}} < \theta < \bar{X} + k\frac{\sigma}{\sqrt{n}}\right) \geq 1 - 1/k^2$$

$$k = 5 \quad 1 - \frac{1}{k^2} = 1 - 1/25$$

$$= 1 - 4/100 = 0.96$$

Central tendency
location

Dispersion
scale

Mean

Variance (squared deviation)

$$\mathrm{Var}[X] = E[(X - E[X])^2]$$

MODE → most common value

RANGE (max/min value)

MEDIAN
↳ middle value

QUARTILES
INTERQUARTILE RANGE
↳ middle 50% of value

Variance

$$\min_{a \in \mathbb{R}} E[(X-a)^2] = E[(X - \underbrace{E[X]}_{\text{mean}})^2]$$

m is the median of X

then

$$\min_{a \in \mathbb{R}} E[|X-a|] = E[|X-m|]$$

median minimizes absolute deviation