

Q1 Why do you need to apply resampling? Explain it.

Ans: ~~The classical approach~~
We need to apply resampling mainly because, in many real problems, classical theoretical formulae are not reliable or not available.

(1) unknown population distribution:

- many statistical methods assume normality or other specific distributions.
- In practice, we often do not know the true distribution.
- Resampling estimates the sampling distribution directly from the data, without strong assumptions.

(2)

Small sample size:

- with small samples, asymptotic (large-sample) results may be inaccurate.
- Resampling provides better estimates of variability, bias and confidence intervals when data are limited.

(3) Complex statistics :

→ For statistics like medians, percentiles, regression coefficient, or machine learning metrics, exact sampling distributions are difficult or impossible to derive.

→ Resampling gives an empirical distribution of these statistics.

(4) Estimate variability and uncertainty :

→ Resampling helps estimate :

(i) standard error

(ii) bias

(iii) confidence intervals

(5) Fewer assumptions :

→ Traditional methods depend on assumptions.

(normality, equal variance, independence)

→ Resampling is often non-parametric, so it works well when assumptions are violated.

Q. 2 Consider the following set of observations
 $\{4, 5, 6, 7, 8\}$, calculate the jackknife mean.

Also show that jackknife mean is an unbiased estimator of the population mean.

Sol:

Given observations,

$$\{4, 5, 6, 7, 8\}, n=5$$

$$\hat{\theta}_5(x) = \bar{x} = \frac{1}{5} \left(\sum_{i=1}^5 x_i \right) = \frac{1}{5} (4+5+6+7+8)$$

$$\left(\bar{x} - \left(\bar{x} \sum_{i=1}^n \frac{1}{n} \right) \right) \sum_{i=1}^n \frac{1}{(1-n)} = 0$$

The five delete one values are

$$\hat{\theta}_4[x_{(1)}] = \bar{x}_1 = \frac{1}{4} (5+6+7+8)$$

$$= \frac{6.5}{1} = 6.5$$

$$\hat{\theta}_4[x_{(2)}] = \bar{x}_2 = \frac{1}{4} (4+6+7+8)$$

$$= 6.25$$

$$\hat{\theta}_4[x_{(3)}] = \bar{x}_3 = \frac{1}{4} (4+5+7+8)$$

$$= 6$$

$$\hat{\theta}_4[x_{(4)}] = \bar{x}_4 = \frac{1}{4} (4+5+6+8)$$

$$= 5.75$$

$$\hat{\theta}_4[x_{(5)}] = \bar{x}_5 = \frac{1}{4} (4+5+6+7)$$

$$= 5.5$$

∴ The jackknife mean,

$$\bar{\theta}_{jack} = \frac{6.5 + 6.25 + 6 + 5.75 + 5.5}{5} = 6$$

∴ Jackknife mean = 6

$$\left[\begin{array}{l} \text{Here, } \bar{\theta}_j = \bar{\theta} \\ E(\bar{\theta}_j) = E(\bar{\theta}) \\ \Rightarrow E(\bar{\theta}_j) = \mu \end{array} \right]$$

Jackknife mean is an unbiased estimate of the population mean.

$$\bar{x}_{jack} = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{j \neq i} x_j \right)$$

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

$$= \frac{1}{n(n-1)} \left[n \cdot \sum_{j=1}^n x_j - \sum_{i=1}^n x_i \right]$$

$$= \frac{(n-1) \sum_{j=1}^n x_j}{n(n-1)}$$

$$= \frac{1}{n} \sum_{j=1}^n x_j$$

$$= \bar{x}$$

$$\therefore \bar{x}_{jack} = \bar{x}$$

Jackknife mean is an unbiased estimate of the population mean.

T-1
Q ③ Define the Cramér-Rao Inequality.

Sol:

Cramér-Rao Inequality,

$$\text{var}(t) \geq \frac{[T'(\theta)]^2}{E\left[\frac{\partial}{\partial \theta} \log L\right]^2}$$

Proof:

If the frequency function of the continuous or discrete population is $f(x|\theta)$.

The likelihood function of a sample n independent observations $(x_1, x_2, \dots, x_n)$ is

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

We shall simply write this L

L is the joint frequency function of the observations

$$\int \dots \int L \, dx_1 \dots dx_n = 1 \quad (1)$$

Differentiate both side of (1) w.r to θ

$$\int \dots \int \frac{\partial}{\partial \theta} L \, dx_1 \dots dx_n = 0$$

which we may rewrite,

$$E\left(\frac{\delta}{\delta\theta} \log L\right) = \int \dots \int \left(\frac{1}{L} \frac{\delta L}{\delta\theta} \right) L \, dx_1 \dots dx_n = 0$$

which becomes,

$$\int \dots \int \left\{ \left(\frac{1}{L} \frac{\delta L}{\delta\theta} \right)^2 + \frac{\delta^2 \log L}{\delta\theta^2} \right\} L \, dx_1 \dots dx_n = 0$$

on,

$$E\left\{ \left(\frac{\delta \log L}{\delta\theta} \right)^2 \right\} = -E\left(\frac{\delta^2 \log L}{\delta\theta^2} \right)$$

$$E(L) = \int \dots \int L \, dx_1 \dots dx_n = T(\theta) \quad (2)$$

Differentiate (2)

$$\int \dots \int \left(L + \frac{\delta \log L}{\delta\theta} L \right) dx_1 \dots dx_n = T'(\theta)$$

which we may rewrite

$$T'(\theta) = \int \dots \int \left(L + \frac{\delta \log L}{\delta\theta} L \right) dx_1 \dots dx_n$$

By using Cauchy-Schwarz inequality —
we have,

$$\{T'(\theta)\}^2 \leq \int \dots \int \left(L + \frac{\delta \log L}{\delta\theta} L \right)^2 dx_1 \dots dx_n$$

which on the arrangement, becomes, -

$$\text{var}(t) = E\{1 - \tau(\theta)\}^2 \geq \frac{[\tau(\theta)]^2}{E\left[\frac{\partial}{\partial \theta} \log L\right]^2}$$

$$\therefore \text{var}(t) \geq \frac{[\tau(\theta)]^2}{E\left[\frac{\partial}{\partial \theta} \log L\right]^2}$$

Q-1
Q4: How can you find Cramér-Rao Inequality from Bhattachayya lower bound?

Sol:

Bhattachayya lower bound,
 we have,

$$\text{var}(t_n) \geq \sum_{n=1}^s \sum_{p=1}^s \tau(\theta)^{(p)} J_{np}^{-1} \tau(\theta)^{(n)}$$

for $s=1$,

$$\text{var}(t_n) \geq \tau(\theta)^{(1)} J_{11}^{-1} \tau(\theta)^{(1)}$$

$$\text{var}(t_n) \geq \frac{[\tau(\theta)^{(1)}]^2}{J_{11}}$$

Here,

$$J_{11} = E\left[\frac{L^{(1)}}{L} \cdot \frac{L^{(1)}}{L}\right]$$

$$= E\left[\frac{\partial \log L}{\partial \theta}\right]^2$$

$$\text{So, } \text{var}(t_n) \geq \frac{[\tau(\theta)^{(1)}]^2}{J_{11}}$$

$$\text{Var}(\ln) \geq \frac{[E\tau(0)]^2}{E\left[\frac{\text{sigL}}{80}\right]}$$

This is the Cramér-Rao Inequality.

Jackknife and Bootstrapping

22

Qsn: Given the data set: 10, 15, 20, 25, 30, use the Jackknife method to estimate the variance. Test whether the jackknife estimator of variance is unbiased based on this data. Also, construct a 95% confidence interval for the jackknife estimator of variance.

Solution:

Given the data $x = \{10, 15, 20, 25, 30\}$

$$n = 5$$

$$\therefore \text{sample mean, } \bar{x} = \frac{1}{5}(10+15+20+25+30) = 20$$

Leave-one-out samples and their means:

$$\hat{\theta}_4[x_{(1)}] = \bar{x}_4 = \frac{1}{4}(15+20+25+30) = 22.5$$

$$\hat{\theta}_4[x_{(2)}] = \bar{x}_2 = \frac{1}{4}(10+20+25+30) = 21.25$$

$$\hat{\theta}_4[x_{(3)}] = \bar{x}_3 = \frac{1}{4}(10+15+25+30) = 20$$

$$\hat{\theta}_4[x_{(4)}] = \bar{x}_4 = \frac{1}{4}(10+15+20+30) = 18.75$$

$$\hat{\theta}_4[x_{(5)}] = \bar{x}_5 = \frac{1}{4}(10+15+20+25) = 17.5$$

∴ The jackknife mean of leave-one-out means

$$\bar{x}_{(j)} = \frac{22.5 + 21.25 + 20 + 18.75 + 17.5}{5}$$

$$= 20$$

Jackknife estimate of variance,

$$\hat{\sigma}_{jack}^2 = \frac{n-1}{n} \sum_{i=1}^n (\bar{x}_{(i)} - \bar{x}_{(j)})^2$$

$$= \frac{5-1}{5} [(22.5-20)^2 + (21.25-20)^2 + (20-20)^2 + (18.75-20)^2 + (17.5-20)^2]$$

$$= \frac{4}{5} [6.25 + 1.5625 + 0 + 3.75 + 6.25]$$

$$\hat{\sigma}_{jack}^2 = 12.5$$

The sample mean, $\bar{x} = 20$

∴ population variance,

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \frac{1}{5} [(10-20)^2 + (15-20)^2 + (20-20)^2 + (25-20)^2 + (30-20)^2]$$

$$= 50$$

$$\sigma^2 = 50$$

∴ variance of the sample mean, $\text{var}(\bar{x})$

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n}$$

$$= \frac{250}{5}$$

$$= 10$$

$$\hat{V}_{\text{jack}} = 12.5$$

$$\hat{V}_{\text{jack}} \neq \text{var}(\bar{x})$$

$$\text{Bias} = E(\hat{V}_{\text{jack}}) - \text{var}(\bar{x}) = 12.5 - 10 = 2.5$$

∴ the Jackknife estimator variance is biased.

95% confidence interval

Jackknife variance,

$$\hat{V}_{\text{jack}} = 12.5$$

$$SE(\hat{V}_{\text{jack}}) = \sqrt{12.5}$$

$$= 3.536$$

95% confidence interval

$$\hat{V}_{\text{jack}} \pm Z_{0.05} \times SE$$

$$12.5 \pm 1.96 \times (3.536)$$

$$\text{Lower limit} = 12.5 - 1.96(3.536)$$

$$= 5.57$$

$$\text{Upper limit} = 12.5 + 1.96(3.536)$$

$$= 19.43$$

$$\therefore (5.57, 19.43)$$

So, 95% confidence interval for the jackknife estimate of variance

$$5.57 \leq \text{var} \leq 19.43$$

N.B: $\bar{x}_{\text{jack}}$ is the jackknife estimate of the mean and its 95% confidence interval is given by

$$\bar{x}_{\text{jack}} = 20$$

$$\text{Sample variance, } s^2 = \frac{250}{n-1} = 62.5$$

Leave-one-out mean

$$22.5, 21.25, 20, 18.75, 17.5$$

$$\bar{x}_{\text{jack}} = 20$$

$$\hat{\Delta}_{\text{jack}} = 12.5$$

$$SE(\hat{\Delta}_{\text{jack}}) = 3.536$$

confidence interval of jackknife mean,

$$\bar{x}_j \pm t_{\alpha/2, (n-1)} \times SE$$

$$= 20 \pm (t_{0.025, 4}) \times 3.536$$

$$= 20 \pm (2.776 \times 3.536)$$

$$= 20 \pm 9.84$$

$$(10.18, 29.82)$$

N.B: why t-distribution used here?

population mean μ ✓

variance is estimate from data

sample size small (5)

two potential problems are (1) $\bar{x}_j$ is a consistent estimator but $\bar{x}_j$ is not unbiased (2) $\bar{x}_j$ is not efficient (3) $\bar{x}_j$ is not robust (4) $\bar{x}_j$ is not invariant to outliers (5) $\bar{x}_j$ is not symmetric (6) $\bar{x}_j$ is not normal (7) $\bar{x}_j$ is not stable (8) $\bar{x}_j$ is not smooth (9) $\bar{x}_j$ is not differentiable (10) $\bar{x}_j$ is not continuous (11) $\bar{x}_j$ is not bounded (12) $\bar{x}_j$ is not non-negative (13) $\bar{x}_j$ is not non-positive (14) $\bar{x}_j$ is not non-zero (15) $\bar{x}_j$ is not non-infinite (16) $\bar{x}_j$ is not non-finite (17) $\bar{x}_j$ is not non-integer (18) $\bar{x}_j$ is not non-rational (19) $\bar{x}_j$ is not non-irrational (20) $\bar{x}_j$ is not non-real (21) $\bar{x}_j$ is not non-complex (22) $\bar{x}_j$ is not non-imaginary (23) $\bar{x}_j$ is not non-real (24) $\bar{x}_j$ is not non-complex (25) $\bar{x}_j$ is not non-imaginary (26) $\bar{x}_j$ is not non-real (27) $\bar{x}_j$ is not non-complex (28) $\bar{x}_j$ is not non-imaginary (29) $\bar{x}_j$ is not non-real (30) $\bar{x}_j$ is not non-complex (31) $\bar{x}_j$ is not non-imaginary (32) $\bar{x}_j$ is not non-real (33) $\bar{x}_j$ is not non-complex (34) $\bar{x}_j$ is not non-imaginary (35) $\bar{x}_j$ is not non-real (36) $\bar{x}_j$ is not non-complex (37) $\bar{x}_j$ is not non-imaginary (38) $\bar{x}_j$ is not non-real (39) $\bar{x}_j$ is not non-complex (40) $\bar{x}_j$ is not non-imaginary (41) $\bar{x}_j$ is not non-real (42) $\bar{x}_j$ is not non-complex (43) $\bar{x}_j$ is not non-imaginary (44) $\bar{x}_j$ is not non-real (45) $\bar{x}_j$ is not non-complex (46) $\bar{x}_j$ is not non-imaginary (47) $\bar{x}_j$ is not non-real (48) $\bar{x}_j$ is not non-complex (49) $\bar{x}_j$ is not non-imaginary (50) $\bar{x}_j$ is not non-real (51) $\bar{x}_j$ is not non-complex (52) $\bar{x}_j$ is not non-imaginary (53) $\bar{x}_j$ is not non-real (54) $\bar{x}_j$ is not non-complex (55) $\bar{x}_j$ is not non-imaginary (56) $\bar{x}_j$ is not non-real (57) $\bar{x}_j$ is not non-complex (58) $\bar{x}_j$ is not non-imaginary (59) $\bar{x}_j$ is not non-real (60) $\bar{x}_j$ is not non-complex (61) $\bar{x}_j$ is not non-imaginary (62) $\bar{x}_j$ is not non-real (63) $\bar{x}_j$ is not non-complex (64) $\bar{x}_j$ is not non-imaginary (65) $\bar{x}_j$ is not non-real (66) $\bar{x}_j$ is not non-complex (67) $\bar{x}_j$ is not non-imaginary (68) $\bar{x}_j$ is not non-real (69) $\bar{x}_j$ is not non-complex (70) $\bar{x}_j$ is not non-imaginary (71) $\bar{x}_j$ is not non-real (72) $\bar{x}_j$ is not non-complex (73) $\bar{x}_j$ is not non-imaginary (74) $\bar{x}_j$ is not non-real (75) $\bar{x}_j$ is not non-complex (76) $\bar{x}_j$ is not non-imaginary (77) $\bar{x}_j$ is not non-real (78) $\bar{x}_j$ is not non-complex (79) $\bar{x}_j$ is not non-imaginary (80) $\bar{x}_j$ is not non-real (81) $\bar{x}_j$ is not non-complex (82) $\bar{x}_j$ is not non-imaginary (83) $\bar{x}_j$ is not non-real (84) $\bar{x}_j$ is not non-complex (85) $\bar{x}_j$ is not non-imaginary (86) $\bar{x}_j$ is not non-real (87) $\bar{x}_j$ is not non-complex (88) $\bar{x}_j$ is not non-imaginary (89) $\bar{x}_j$ is not non-real (90) $\bar{x}_j$ is not non-complex (91) $\bar{x}_j$ is not non-imaginary (92) $\bar{x}_j$ is not non-real (93) $\bar{x}_j$ is not non-complex (94) $\bar{x}_j$ is not non-imaginary (95) $\bar{x}_j$ is not non-real (96) $\bar{x}_j$ is not non-complex (97) $\bar{x}_j$ is not non-imaginary (98) $\bar{x}_j$ is not non-real (99) $\bar{x}_j$ is not non-complex (100) $\bar{x}_j$ is not non-imaginary

to ensure that the method is unbiased

bootstrapping is a resampling method

bootstrapping is a resampling method

bootstrapping is a resampling method

bootstrapping is a resampling method

bootstrapping is a resampling method

Q What is Jackknife estimation method?

Ans: The Jackknife is a resampling technique used to estimate the bias, standard error and CI of the statistics.

It is used when:-

- (1) Sample size is small.
- (2) No parameter assumptions can be made.
- (3) We want to reduce the effects of outliers. It is a computational method that involves systematically leaving out one observation at a time from our sample and recalculating the statistic.

Let, $s^2 = \frac{\sum (x_i - \bar{x})^2}{N}$ is a consistent estimator but biased estimator of σ^2 , then we can add a simple adjustment factor to make an unbiased estimator.

Maurice Quenette gives a method of overcoming this type of biasness. This method is called Jackknife estimation method.

$$\begin{aligned} E(s^2) &= \mu_2 - \frac{1}{n} \mu_2 \\ E(s^2) &= \mu_2 - \frac{1}{n} \mu_2 + \frac{1}{n} \mu_2 \\ E(s^2) &= \mu_2 \end{aligned}$$

23 Qsn: you have a dataset $x = \{x_1, x_2, \dots, x_n\}$ and the parameter of interest is the sample mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

(i) Derive the jackknife estimate of bias of $\bar{x}$.

Sol:

For the i -th jackknife sample, remove x_i

The leave-one-out mean is

$$\bar{x}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} x_j = \frac{n\bar{x} - x_i}{n-1}$$

The mean of Jackknife estimate,

$$\begin{aligned}\bar{x}_{\text{jack}} &= \frac{1}{n} \sum_{i=1}^n \bar{x}_{(i)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{n\bar{x} - x_i}{n-1} \\ &= \frac{1}{n(n-1)} \left(n^2 \bar{x} - \sum_{i=1}^n x_i \right)\end{aligned}$$

$$\begin{aligned}\bar{x}_{\text{jack}} &= \frac{1}{n(n-1)} (n^2 \bar{x} - n\bar{x}) \\ &= \bar{x}\end{aligned}$$

$$\therefore \bar{x}_{\text{jack}} = \bar{x}$$

The jackknife bias estimator is

$$\widehat{\text{Bias}}_{\text{jack}} = (n-1) (\bar{x}_{\text{jack}} - \bar{x})$$

$\therefore$ This is not bias.

The sample mean is an unbiased estimator of μ .

22

[E]

Bootstrap sampling?

Bootstrap sampling:

$\Rightarrow$ suppose we have observed an independent sample

$x = (x_1, x_2, \dots, x_n)$ from which we can compute a statistic of interest $s(x) = \{\bar{x}, \text{var}(x)\}$.

A bootstrap sample $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ is obtained by random sampling n times with replacement from the original sample $(x_1, \dots, x_n)$.

Let $n=7$

we may obtain

$$x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

$$x = (x_3, x_4, x_3, x_7, x_4, x_3, x_1)$$

The bootstrap process is -

$$x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7) \rightarrow \text{original sample}$$

$$x_1^*, x_2^*, \dots, x_n^* \rightarrow \text{bootstrap sample}$$

$$s(x_1^*), s(x_2^*), \dots, s(x_n^*) \rightarrow \text{bootstrap statistics}$$

22

Describe the Bootstrap algorithm for estimating standard errors.

Solution:

The bootstrap algorithm for estimating SE.

- (1) select B independent bootstrap samples $x^{*1}, x^{*2}, \dots, x^{*B}$ each consisting of n data values drawn with replacement from x .
- (2) Evaluate the bootstrap replication $s(x^{*1}), s(x^{*2}), \dots, s(x^{*B})$ obtained by the value of the statistics $s(x)$ on each bootstrap sample where,

$$\hat{\theta}(b) = s(x^{*b}), \quad b = 1, 2, \dots, B$$

- (3) Estimate the $se_F(\hat{\theta})$ by the sample standard deviation of the B replication

$$se(\hat{\theta}) = \hat{se}_{boot} = \left\{ \frac{\sum_{b=1}^B [\hat{\theta}^{*}(b) - \hat{\theta}^{*}(\cdot)]^2}{B-1} \right\}^{1/2}$$

where, $\hat{\theta}^{*}(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^{*}(b)}{B}$

Q Bootstrap estimate of bias

Ans: Suppose we want to estimate a real valued parameter $\theta = t(F)$ and the plug in estimate $\hat{\theta} = t(\hat{F})$

Then bias of $\hat{\theta} = t(\hat{F})$ as an estimate of θ is defined to be the difference between the expectation of $\hat{\theta}$ and real valued parameter θ .

Then, $\text{bias}_F(\hat{\theta}, \theta) = E_F[\hat{\theta}] - t(F)$

The bootstrap estimate of bias based on B replication

$$\hat{\text{bias}}_B = \hat{\theta}^* - t(\hat{F})$$

where, $t(\hat{F})$ is an plug-in estimate of θ and may differ from $\hat{\theta} = t(\hat{F})$

Q Question: Chapman-Robins and Kiefer Inequality.

Ans: CRK gives a lower bound for the variance of an estimate but doesn't require regularity conditions of the Cramér-Rao inequality.

Theorem: Let $x_1, x_2, \dots, x_n$ be a random variables drawn from the probability density function $f(x; \theta)$; $\theta \in \Xi$. Again, $x_1, x_2, \dots, x_n$ be a random variables with probability density function $f(x; \phi)$; $\phi \in \Xi$. Also, let T be an unbiased estimator of both $\tau(\theta)$ and $\tau(\phi)$ with $E[T^2] < \infty$, $\forall \theta, \phi \in \Xi$. If $\theta \neq \phi$, assume that $f(x; \theta)$ and $f(x; \phi)$ are different.

$$\text{Also, } S(\theta) \supset S(\phi)$$

$$\Rightarrow \{f(x_i; \theta) > 0\} \supset \{f(x_i; \phi) > 0\}$$

Then,

$$\text{var}[T] \geq \frac{[\tau(\phi) - \tau(\theta)]^2}{\text{var}\left[\frac{f(x_i; \phi)}{f(x_i; \theta)}\right]}$$

Proof:

Since, T is an unbiased estimator of $\tau(\phi)$, so we can write —

$$E[T(x)] = \tau(\phi) \quad ; \quad \forall \phi \in \Xi$$

Hence, $\phi \neq \theta$

$$\int_{S(\theta)}^{\tau(\phi)} \frac{f(x; \phi) - f(x; \theta)}{f(x; \theta)} f(x; \theta) dx$$

$= T(\phi) - T(\theta)$; which yields.

$$\text{cov} \left\{ T(x), \frac{f(x; \phi)}{f(x; \theta)} - 1 \right\}$$

Now, by applying Cauchy-Schwarz inequality, we get

$$\text{cov} \left\{ T(x), \left[\frac{f(x; \phi)}{f(x; \theta)} - 1 \right] \right\}^2 \leq \text{var}[T(x)] \cdot \text{var} \left[\frac{f(x; \phi)}{f(x; \theta)} - 1 \right]$$

$$\Rightarrow \{ T(\phi) - T(\theta) \}^2 \leq \text{var}[T(x)] \cdot \text{var} \left[\frac{f(x; \phi)}{f(x; \theta)} - 1 \right]$$

$$\Rightarrow \text{var}[T(x)] \geq \frac{\{ T(\phi) - T(\theta) \}^2}{\text{var} \left[\frac{f(x; \phi)}{f(x; \theta)} - 1 \right]}$$

CRLB inequality for lower bound is more accurate than Cramér-Rao inequality, and there is no

Pitman closer and robust :

22

Qsn:

What do you mean by pitman closer and pitman closer Estimator?

Ans:

Pitman closer and pitman closer :

Let, $T = t(x_1, \dots, x_n)$ and $T' = t'(x_1, \dots, x_n)$ be two estimators of $\tau(\theta)$. T' is called a pitman closer estimator of $\tau(\theta)$ than T if and only if,

$$P_\theta [|T' - \tau(\theta)| < |T - \tau(\theta)|] \geq \frac{1}{2}$$

for each θ in Θ

An estimator T^* is called pitman closer if it is pitman closer than any other estimator.

$$\therefore \text{var}(\hat{\theta}) \geq \frac{1}{\sum_{i=1}^n I(\hat{\theta})}$$

↓
Cramér-Rao lower bound.



Location invariant :

An estimator $T = t(x_1, \dots, x_n)$ is defined to be location invariant if and only if

$$t(x_1+c, x_2+c, \dots, x_n+c) = t(x_1, \dots, x_n) + c = T + c$$

for all values $(x_1, x_2, \dots, x_n)$ and all c .

Scale invariant

An estimator $T = t(x_1, \dots, x_n)$ is defined to be scale invariant if and only if —

$$t(cx_1, cx_2, \dots, cx_n) = t(x_1, x_2, \dots, x_n)$$

for all values $(x_1, x_2, \dots, x_n)$ and c .

Q

Pitman estimator for location:

Let $x_1, x_2, \dots, x_n$ denote a random sample from the density $f(x; \theta)$, where θ is a location parameter then the estimator —

$$t(x_1, \dots, x_n) = \frac{\int \theta \prod_{i=1}^n f(x_i; \theta) d\theta}{\int \prod_{i=1}^n f(x_i; \theta) d\theta}$$

is the estimator of θ which has uniformly smallest mean squared error or smallest risk estimator within the class of location invariant estimator.

Ex pitman estimator for scale:

Let $x_1, \dots, x_n$ be a random sample from the density $f(\cdot; \theta)$, where $\theta > 0$ is a scale parameter. Assume that $f(x; \theta) = 0$ for $x \leq 0$. That is the random variable x_i assume only the positive values. Then the estimators,

$$t(x_1, \dots, x_n) = \frac{\int_0^\infty \frac{1}{\theta^n} \prod_{i=1}^n f(x_i; \theta) d\theta}{\int_0^\infty \frac{1}{\theta^3} \prod_{i=1}^n f(x_i; \theta) d\theta}$$

is the estimator of θ which has uniformly smallest risk within the class of scale invariant estimators for the loss function

$$L(t; \theta) = \frac{(t - \theta)^2}{\theta^2}$$

22
Ex Q: Let $x_1, x_2, \dots, x_n$ be a random sample of size n from the density.

$$f(x; \theta) = \frac{1}{\theta}, \quad 0 < x < 2\theta$$

Find the pitman estimator for the scale parameter.

Sol:

$$t(x_1, \dots, x_n) = \frac{\int_{Y_n} \frac{1}{\theta^n} \prod_{i=1}^n \frac{1}{\theta} d\theta}{\int_{Y_n} \frac{1}{\theta^3} \prod_{i=1}^n \frac{1}{\theta} d\theta}$$

$$= \frac{\int_{Y_n} \theta^{-2} \cdot \theta^{-n} d\theta}{\int_{Y_n} \theta^{-3} \cdot \theta^{-n} d\theta}$$

$$= \frac{\int_{Y_n} \theta^{-(n+2)} d\theta}{\int_{Y_n} \theta^{-(n+3)} d\theta}$$

$$= \frac{\left[\frac{\theta^{-(n+2)+1}}{-(n+2)+1} \right]_{Y_n}^{\infty}}{\left[\frac{\theta^{-(n+3)+1}}{-(n+3)+1} \right]_{Y_n}^{\infty}}$$

$$= \frac{\left[\frac{\theta^{-(n+1)}}{-(n+1)} \right]_{Y_n}^{\infty}}{\left[\frac{\theta^{-(n+2)}}{-(n+2)} \right]_{Y_n}^{\infty}}$$

$$= \frac{n+2}{n+1} \cdot \frac{\left[\theta^{-(n+1)} \right]_{Y_n}^{\infty}}{\left[\theta^{-(n+2)} \right]_{Y_n}^{\infty}}$$

$$= \frac{n+2}{n+1} \cdot \frac{\theta^{-n-1}}{\theta^{-n-2}}$$

$$= \frac{n+2}{n+1} \cdot \theta^{-n-1+1}$$

$$= \frac{n+2}{n+1} \cdot \theta$$

$$\therefore \frac{n+1}{n+1} \cdot Y_n.$$

Ex Example 10

Let $x_1, \dots, x_n$ be a random sample from a uniform distribution over the interval $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Find the Pitman estimator of θ . θ is in location parameter.

Sol:

$$t(x_1, \dots, x_n) = \frac{\int \theta \prod_{i=1}^n f(x_i; \theta) d\theta}{\int \prod_{i=1}^n f(x_i; \theta) d\theta}$$

$$\int \theta \cdot \prod_{i=1}^n I(\theta - \frac{1}{2}, \theta + \frac{1}{2})(x) d\theta$$

$$\int \prod_{i=1}^n I(\theta - \frac{1}{2}, \theta + \frac{1}{2})(x) d\theta$$

$$\int \theta \cdot \prod_{i=1}^n I(x_i - \frac{1}{2}, x_i + \frac{1}{2})(\theta) d\theta$$

$$\int \prod_{i=1}^n I(x_i - \frac{1}{2}, x_i + \frac{1}{2})(\theta) d\theta$$

$$\int_{x_{n-\frac{1}{2}}}^{x_{1+\frac{1}{2}}} \theta d\theta$$

$$\int_{x_{n-\frac{1}{2}}}^{x_{1+\frac{1}{2}}} d\theta$$

$$= \prod_{i=1}^n \frac{1}{2} \left[\frac{\theta^2}{2} \right]_{x_{n-\frac{1}{2}}}^{x_{1+\frac{1}{2}}}$$

$$= \prod_{i=1}^n \frac{1}{2} \left[\theta \right]_{x_{n-\frac{1}{2}}}^{x_{1+\frac{1}{2}}}$$

$$= \frac{1}{2} \frac{(Y_1 + \frac{1}{2})^2 - (Y_n - \frac{1}{2})^2}{(Y_1 + \frac{1}{2}) - (Y_n - \frac{1}{2})}$$

$$= \frac{1}{2} \frac{((Y_1 + \frac{1}{2}) + (Y_n - \frac{1}{2}))((Y_1 + \frac{1}{2}) - (Y_n - \frac{1}{2}))}{((Y_1 + \frac{1}{2}) - (Y_n - \frac{1}{2}))}$$

$$= \frac{1}{2} (Y_1 + \frac{1}{2} + Y_n - \frac{1}{2})$$

$$= \frac{Y_1 + Y_n}{2}$$

$$\therefore \frac{Y_1 + Y_n}{2}$$

Ex Example :- Let $x_1, \dots, x_n$ be a random sample from a density $f(x; \lambda) = \frac{1}{\lambda} \exp(-\frac{x}{\lambda}) I_{(0, \infty)}(x)$. Find the pitman estimator of λ for the scale parameter.

Sol:

$$t(x_1, \dots, x_n) = \frac{\int_0^\infty \frac{1}{\lambda^2} \prod_{i=1}^n f(x_i; \lambda) d\lambda}{\int_0^\infty \frac{1}{\lambda^3} \prod_{i=1}^n f(x_i; \lambda) d\lambda}$$

$$= \frac{\int_0^\infty \frac{1}{\lambda^2} \prod_{i=1}^n \frac{1}{\lambda} \exp(-\frac{x_i}{\lambda}) d\lambda}{\int_0^\infty \frac{1}{\lambda^3} \prod_{i=1}^n \frac{1}{\lambda} \exp(-\frac{x_i}{\lambda}) d\lambda}$$

$$= \frac{\int_0^\infty \frac{1}{\lambda^{n+2}} \prod_{i=1}^n \exp(-\frac{x_i}{\lambda}) d\lambda}{\int_0^\infty \frac{1}{\lambda^{n+3}} \prod_{i=1}^n \exp(-\frac{x_i}{\lambda}) d\lambda}$$

$$= \frac{\int_0^\infty \frac{1}{\lambda^2} \cdot \frac{1}{\lambda^n} e^{-\frac{\sum x_i}{\lambda}} d\lambda}{\int_0^\infty \frac{1}{\lambda^3} \cdot \frac{1}{\lambda^n} e^{-\frac{\sum x_i}{\lambda}} d\lambda}$$

$$= \frac{\int_0^\infty \frac{1}{\lambda^{n+2}} e^{-\frac{\sum x_i}{\lambda}} d\lambda}{\int_0^\infty \frac{1}{\lambda^{n+3}} e^{-\frac{\sum x_i}{\lambda}} d\lambda}$$

Let,

$$\frac{\sum x_i}{\lambda} = z$$

$$\Rightarrow \lambda = \frac{\sum x_i}{z}$$

$$\Rightarrow d\lambda = -\frac{\sum x_i}{z^2} dz$$

Now,

$$\therefore f(x_1, x_2, \dots, x_n) = \frac{\int_0^\infty \left(\frac{\sum x_i}{z}\right)^{-(n+2)} e^{-z} \left(-\frac{\sum x_i}{z^2}\right) dz}{\int_0^\infty \left(\frac{\sum x_i}{z}\right)^{-(n+3)} e^{-z} \left(-\frac{\sum x_i}{z^2}\right) dz}$$

$$= (\sum x_i)^{-n-2+n+3} \cdot \frac{\int_0^\infty z^{n+2-2} e^{-z} dz}{\int_0^\infty z^{n+3-2} e^{-z} dz}$$

$$= \sum x_i \cdot \frac{\int_0^\infty z^n e^{-z} dz}{\int_0^\infty z^{n+1} e^{-z} dz}$$

$$= \sum x_i \cdot \frac{\Gamma(n+1)}{\Gamma(n+2)}$$

$$= \sum x_i \cdot \frac{n!}{(n+1)!}$$

$$= \sum x_i \frac{n!}{(n+1)!}$$

$$= \sum x_i \cdot \frac{n!}{(n+1)n!}$$

$$= \frac{\sum x_i}{n+1}$$

So, it is pitman scale estimator.

Q Robust statistics?

Ans: Robust statistical methods try to fit the model imposed by the majority of the data. They aim to find a 'robust' fit, which is similar to the fit we would have found without the outliers.

**** When we apply robust Estimators?**

Ans: In practice, we make many assumptions in practice inference and any one or all of these may violate.

Thus, we may seldom have accurate knowledge about the underlying distribution. Similarly, the assumptions of event iid may not hold. Any test or estimators that perform well under modification or under assumptions is usually referred to as robust estimator.

to the test. A test function T_n is

defined as $T_n = T_n(X_1, \dots, X_n)$

such that $T_n \rightarrow T$ as $n \rightarrow \infty$

where T is a function of the distribution

▣ Generalised likelihood Ratio Test:

Let, $L(\theta; x_1, \dots, x_n)$ be the likelihood function for a sample $x_1, \dots, x_n$ having joint density function $f_{x_1, \dots, x_n}(x_1, \dots, x_n; \theta)$, where $\theta \in \Theta$. The generalized likelihood ratio, denoted by λ or λ_n is defined to be —

$$\lambda = \lambda_n = \lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta; x_1, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)}$$

Now that, λ is a function of $x_1, \dots, x_n$ namely $\lambda(x_1, \dots, x_n)$. When the observations are replaced by their corresponding random variables

$x_1, \dots, x_n$, then we write Λ for λ , that is $\Lambda = \lambda(x_1, \dots, x_n)$. Λ is a function of the random variables $x_1, \dots, x_n$, and is itself a random variable. In fact, Λ is a statistic since it doesn't depend on unknown parameter.

▣ Most powerful test: A test γ^* of

$H_0: \theta = \theta_0$ versus $H_1: \theta$ is defined to be a most powerful test of size α ($0 < \alpha < 1$) if and only if —

$$(i) \pi_{\gamma^*}(\theta_0) = \alpha$$

$$(ii) \pi_{\gamma^*}(\theta_1) > \pi_{\gamma}(\theta_1) \text{ for any other test } \gamma, \text{ for which } \pi_{\gamma}(\theta_0) \leq \alpha$$

A test γ^* is most powerful of size α if it has size α and if among all other tests of size α or less it has the largest power. or, a test γ^* is most powerful of size α if it has the size of its Type-I error equal to α and has smallest size Type-II error among all other tests with size of Type-I error α or less.

Uniformly Most powerful:

A test γ^* of $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta - \Theta_0$ is defined to be a uniformly most powerful test size α test if and only if —

$$(i) \sup_{\theta \in \Theta_0} \pi_{\gamma^*}(\theta) = \alpha$$

$$(ii) \pi_{\gamma^*}(\theta) \geq \pi_{\gamma}(\theta) \text{ for all } \theta \in \Theta - \Theta_0 \text{ and for any test } \gamma \text{ with size less than or equal to } \alpha.$$

A test γ^* is uniformly most powerful test if

size α if ϕ has size α and if among all tests of size less than or equal to α , it has the largest power for all alternative values of θ . The adjective 'uniformly' refers to "all" alternative θ values.

Neyman Pearson

State: Let $x_1, \dots, x_n$ be a random sample from a distribution that has pdf or pmf $f(x; \theta)$. Then the likelihood of $x_1, \dots, x_n$ is

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta), \text{ for } x = (x_1, \dots, x_n)$$

Let θ' and θ'' be distinct fixed values of θ so that $\Omega = \{ \theta : \theta = \theta', \theta'' \}$ and let K be a positive number.

Let C be a subset of the sample space such that,

(a) $\frac{L(\theta'; x)}{L(\theta''; x)} \leq K$, for each point $x \in C$ then reject H_0

(b) $\frac{L(\theta'; x)}{L(\theta''; x)} > K$, for each point $x \in C$, then accept H_0 .

(c) $\alpha = P_{H_0}[x \in C]$

Then C is the best critical region of size α for testing $H_0: \theta = \theta' \text{ vs } H_1: \theta \neq \theta''$

Derive:

Proof:

Let C be a critical region of size α and A be another critical region of size α . We shall let —

$\int_{\mathcal{R}} L(\theta; x_1, \dots, x_n) dx_1, \dots, dx_n$ be denoted by $\int_{\mathcal{R}} L(\theta)$.

We have to prove that —

$$\int_C L(\theta'') - \int_A L(\theta'') > 0$$

Since C is the union of the disjoint sets $C \cap A$ and $C \cap A^c$ and A is the union of the disjoint sets $A \cap C$ and $A \cap C^c$, we have

$$\begin{aligned} \int_C L(\theta'') - \int_A L(\theta'') &= \int_{C \cap A} L(\theta'') + \int_{C \cap A^c} L(\theta'') - \int_{A \cap C} L(\theta'') - \int_{A \cap C^c} L(\theta'') \\ &= \int_{C \cap A^c} L(\theta'') - \int_{A \cap C^c} L(\theta'') \end{aligned}$$

by the hypothesis of the theorem,

$$L(\theta'') \geq \frac{1}{K} L(\theta') \text{ at each point of } C \text{ and}$$

hence at each point of $C \cap A^c$, thus—

$$\int_{C \cap A^c} L(\theta'') \geq \frac{1}{K} \int_{C \cap A^c} L(\theta')$$

But $L(\theta'') \leq \frac{1}{K} L(\theta')$ at each of point of

$C \cap A^c$ and hence at each point of $A \cap C^c$;

These inequality $\int_{A \cap C^c} L(\theta'') \leq \frac{1}{K} \int_{A \cap C^c} L(\theta')$

These inequality imply that,—

$$\int_{C \cap A^c} L(\theta'') - \int_{A \cap C^c} L(\theta'') \geq \frac{1}{K} \int_{C \cap A^c} L(\theta') - \frac{1}{K} \int_{A \cap C^c} L(\theta')$$

From (i),

$$\int_C L(\theta'') - \int_A L(\theta'') \geq \frac{1}{K} \left[\int_{C \cap A^c} L(\theta') - \int_{A \cap C^c} L(\theta') \right] \quad (ii)$$

However,

$$\int_{C \cap A^c} L(\theta') - \int_{A \cap C^c} L(\theta') = \int_{C \cap A} L(\theta') + \int_{C \cap A^c} L(\theta') - \int_{A \cap C} L(\theta') - \int_{A \cap C^c} L(\theta')$$

$$= \int_C L(\theta') - \int_A L(\theta')$$

$$= \alpha - \alpha$$

$$= 0$$

Now, it is substituted in (ii), we obtain -

$$\int_C L(\theta'') - \int_A L(\theta'') > 0$$

22
 Let $x_1, \dots, x_n$ be a random sample from $N(0, \theta)$.
 Find the uniformly most powerful test (UMP) if it is
 exist for testing hypothesis -

$$H_0: \theta = \theta' \quad \text{and} \quad H_1: \theta > \theta'$$

Sol:

The distribution,

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2} \frac{x^2}{\theta}\right\}$$

The likelihood function,

$$L(\theta; x) = \left(\frac{1}{2\pi\theta}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n x_i^2\right\}$$

To find the uniformly most powerful test (UMP), the Neyman-Pearson Lemma becomes -

$$\frac{L(\theta'; x)}{L(\theta'', x)} \leq K$$

$$\Rightarrow \frac{\left(\frac{1}{2\pi\theta'}\right)^{\frac{n}{2}} \cdot \exp\left\{-\frac{1}{2\theta'} \sum x_i^2\right\}}{\left(\frac{1}{2\pi\theta''}\right)^{\frac{n}{2}} \cdot \exp\left\{-\frac{1}{2\theta''} \sum x_i^2\right\}} \leq K$$

$$\Rightarrow \left(\frac{\theta''}{\theta'}\right)^{\frac{n}{2}} \cdot \exp\left\{-\frac{1}{2\theta'} \sum x_i^2 + \frac{1}{2\theta''} \sum x_i^2\right\} \leq K$$

$$\Rightarrow \left(\frac{\theta''}{\theta'}\right)^{\frac{n}{2}} \cdot \exp\left\{\frac{-\theta'' \sum x_i^2 + \theta' \sum x_i^2}{2\theta'\theta''}\right\} \leq K$$

$$\Rightarrow \left(\frac{\theta''}{\theta'}\right)^{\frac{n}{2}} \cdot \exp\left\{\frac{-\sum x_i^2 (\theta'' - \theta')}{2\theta'\theta''}\right\} \leq K$$

$$\Rightarrow \frac{n}{2} \log\left(\frac{\theta''}{\theta'}\right) - \frac{\sum x_i^2 (\theta'' - \theta')}{2\theta'\theta''} \leq \log K$$

$$\Rightarrow \frac{-\sum x_i^2 (\theta'' - \theta')}{2\theta'\theta''} \leq \log K - \frac{n}{2} \log\left(\frac{\theta''}{\theta'}\right)$$

$$\Rightarrow \sum x_i^2 \geq \left(\frac{n}{2} \log\left(\frac{\theta''}{\theta'}\right) - \log K\right) \frac{2\theta'\theta''}{(\theta'' - \theta')} = c$$

The set $C = \{ (x_1, x_2, \dots, x_n) : \sum x_i^2 > c \}$ is the best critical region for testing $H_0: \theta = \theta'$ vs $H_1: \theta = \theta''$.

If H_0 is true, the random variable $\frac{\sum x_i^2}{\theta'} \sim \chi_{df=n}^2$ since -

$$\alpha = P_{\theta'} \left(\sum_{i=1}^n \frac{x_i^2}{\theta'} > \frac{c}{\theta'} \right)$$

Then, $C = \{ (x_1, x_2, \dots, x_n) : \sum x_i^2 > c \}$ is the best critical region of size α for testing $H_0: \theta = \theta'$ vs $H_1: \theta = \theta''$. Moreover, for each number θ'' greater than θ' , the foregoing argument holds. That is, $C = \{ (x_1, \dots, x_n) : \sum_{i=1}^n x_i^2 > c \}$ is uniformly most critical region of size α for testing $H_0: \theta = \theta'$ vs $H_1: \theta > \theta'$.