

Point Estimation

- When sampling from a population $\sim f(x|\theta)$
knowing $\theta \Leftrightarrow$ knowing the entire population.

\Rightarrow From the observed sample, we want to estimate
the parameter θ .

- Def: Point estimation is $\vec{w}(x)$, i.e. Any statistic
is a point estimation.

(1) Method of

finding estimations:

① Method of moments:

If we want to estimate $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$

Denote $M_k = E(M_k)$. $\hat{M}_k = \sum_i^n \frac{x_i^k}{n}$

Then $M_1 = M_1(\theta_1, \theta_2, \dots, \theta_k)$

$M_2 = M_2(\theta_1, \dots, \theta_k)$

Replace M_k by

\hat{M}_k (estimator)

$M_k = M_k(\theta_1, \dots, \theta_k)$

Solve: $\theta_i = \theta_i(\hat{M}_1, \hat{M}_2, \dots, \hat{M}_k)$ is an estimator!

Remark: A Technique: moment matching:

e.g. (Sinterearthwaste Approximation)

$Y_i \sim \chi^2_{r_i}$, we want to find v.s.t. $\sum_i^k a_i Y_i$

will approximate $\frac{\chi^2}{v}$. But v is unknown.

Note that:
$$\begin{cases} E(\sum a_i Y_i) = \sum a_i r_i = E\left(\frac{\chi^2}{v}\right) = 1 \\ E(\sum a_i Y_i)^2 = E\left(\frac{\chi^2}{v}\right)^2 = \frac{2}{v} + 1 \end{cases}$$

$$\Rightarrow v \approx \frac{2}{(\sum_i^k a_i Y_i)^2 - 1}, \text{ it's complicated and inefficient.}$$

But to cancel the "1":

$$\begin{aligned} E(\sum a_i Y_i)^2 &= \text{Var}(\sum a_i Y_i) + (E(\sum a_i Y_i))^2 \\ &= (E(\sum a_i Y_i))^2 \left(\frac{\text{Var}(\sum a_i Y_i)}{(E(\sum a_i Y_i))^2} + 1 \right) \\ &= \frac{\text{Var}(\sum a_i Y_i)}{(E(\sum a_i Y_i))^2} + 1. \quad \text{since } \text{Var}(D) = \sum_i^k \frac{a_i^2 E(Y_i)^2}{r_i} \end{aligned}$$

$$\therefore \hat{v} = \frac{(E(\sum a_i Y_i))^2}{\sum \frac{a_i^2}{r_i} Y_i^2}$$

② Maximum Likelihood

Estimators:

Def: Likelihood Function: $L(\vec{\theta} | \vec{x}) = L(\theta_1, \dots, \theta_k | x_1, \dots, x_n)$

$= \prod f(x_i | \theta_1, \dots, \theta_k)$. X_i s, i.i.d. samples.

A MLE $\hat{\theta} = \hat{\theta}(\vec{x})$ is that maximum the

Likelihood Func. i.e. $L(\hat{\theta}(\vec{x}) | \vec{x}) = \sup_{\theta \in \Theta} L(\theta | \vec{x})$

Two problems

→ i) How to find MLE?

→ ii) The numerical sensitivity.

For i): Let $\frac{\partial L(\theta | \vec{x})}{\partial \theta} = 0$, check $\frac{\partial^2}{\partial \theta^2} L |_{\hat{\theta}} < 0$

so that it's global maximum. ($Z = \log L(\theta | \vec{x})$)

If θ must be an integer:

$$\text{Calculate: } \frac{L(\theta=k | \vec{x})}{L(\theta=k+1 | \vec{x})} \neq 1$$

Thm. (Invariance property of MLEs)

If $\hat{\theta}$ is MLE of θ , Then for any func. $Z(\cdot)$.

$Z(\hat{\theta})$ is MLE of $Z(\theta)$.

Df. Define the indirect Likelihood Func.

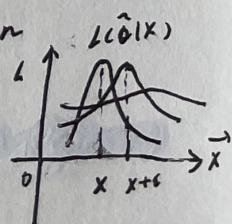
$$L^*(\eta | \vec{x}) = \sup_{\{\theta | Z(\theta)=\eta\}} L(\theta | \vec{x}), L^*(\hat{\eta} | \vec{x}) = \sup_n \sup_{\{\theta | Z(\theta)=\eta\}} L(\theta | \vec{x})$$

$$\text{Then prove: } L^*(\hat{\eta} | \vec{x}) = L^*(Z(\hat{\theta}) | \vec{x})$$

For ii):

We expect that: If base our calculation on $L(\theta | \vec{x} + \vec{\epsilon})$ ($\hat{\theta}_1$ is its MLE). Then $\hat{\theta}_1$ should be close to $\hat{\theta}$, for $\vec{\epsilon}$ is small.

However, this only happens when the likelihood Function is very flat in the neighbour of $f(\vec{x} | \hat{\theta})$ (or $\hat{\theta}=\infty$)



It will change a lot!

③ Bayes Estimators:

Assume $\theta \sim \pi(\theta)$. (It's subjective)

Given $X \sim f(x|\theta)$. \Rightarrow Calculate $m(\theta|\vec{x})$

Then the estimator is $E_{\theta} (m(\theta|\vec{x})) = T(\vec{x})$.

④ The EM algorithm:

- It's designed to find MLEs, specifically, by iteration.

(2) The method of

Evaluating Estimators:

- Now, we face the task of choosing estimators

Actually, this part is part of Decision Theory.

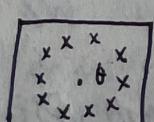
① Mean Square Error:

- Def: MSE of estimator W of para. θ is

$$E_{\theta} (W - \theta)^2.$$

Actually, $E_{\theta} (W - \theta)^2 = \text{Var}_{\theta}(W) + (E_{\theta}(W) - \theta)^2$

$$\triangleq \text{Var}_{\theta}(W) + \text{Bias}_{\theta}(W)^2$$



①
 w_1



②
 w_2

$$\Rightarrow \text{Var}_{\theta}(w_1) > \text{Var}_{\theta}(w_2)$$

$$\text{But } \text{Bias}_{\theta}(w_1) < \text{Bias}_{\theta}(w_2)$$

How to compare MSE?

Remark: MSE can be useful criterion for finding the best estimator in a class of equivariant estimators:

$\left\{ \begin{array}{l} \text{Measure-Equivaria: } W(\vec{x}) \text{ estimates } \theta \Rightarrow \bar{g}(W(\vec{x})) \text{ estimates } \bar{g}(\theta) \\ \text{Formal-Invaria: } W(\vec{x}) \text{ estimates } \theta \Rightarrow W(g(\vec{x})) \text{ estimates } g(\theta) \end{array} \right.$

($\bar{g}: \mathbb{R} \rightarrow \mathbb{R}$, but $g: \mathbb{R} \rightarrow \mathbb{R}$, so it's denoted differently)

\Rightarrow i.e. $\bar{g}(W(\vec{x})) = W(g(\vec{x}))$ estimates $\bar{g}(\theta)$.

② Best Unbiased Estimators:

Note that Best MSE is a large class
So we limit on the class of unbiased estimators

Def: An estimator W^* is called uniform of $Z(\theta)$
minimum variance unbiased estimator (UMVUE)

If any other estimator W , $\text{Var}_\theta(W^*) \leq \text{Var}_\theta(W)$, $\forall \theta$.

where $E_\theta(W) = Z(\theta)$, i.e. $W \in C_z = \{W | E_\theta(W) = Z(\theta)\}$.

\Rightarrow But find an UMVUE isn't a easy task!

Approach one:

If estimate $Z(\theta)$, where $X \sim f(x|\theta)$, we can specify a lower bound $B(\theta)$ on the variance of

any unbiased estimator of θ . Then we try to find W^* s.t. $\text{Var}_\theta(W^*) = \text{B}(\theta)$.

Thm. C (Cramér-Rao Inequality)

$\vec{x} = (x_1, \dots, x_n) \sim f(\vec{x}|\theta)$. $W(\vec{x})$ an estimator satisfying regular condition: $\frac{\lambda}{\lambda\theta} E_\theta W(\vec{x}) = \int_{\vec{x}} \frac{\partial}{\partial \theta} [W(\vec{x}) f(\vec{x}|\theta)] d\vec{x}$.

And $\text{Var}_\theta(W(\vec{x})) < \infty$.

$$\text{Then: } \text{Var}_\theta(W(\vec{x})) \geq \frac{\left(\frac{\lambda}{\lambda\theta} E_\theta (W(\vec{x})) \right)^2}{E_\theta \left(\frac{\partial}{\partial \theta} \log f(\vec{x}|\theta) \right)^2}$$

Pf. Lemma. i) An identity:

$$\text{Since } \int_{\vec{x}} f(\vec{x}|\theta) d\vec{x} = 1. \therefore \frac{\partial}{\partial \theta} \int_{\vec{x}} f(\vec{x}|\theta) d\vec{x} = 0$$

$$\text{LHS} = \int_{\vec{x}} \frac{\partial}{\partial \theta} f(\vec{x}|\theta) d\vec{x} = \int_{\vec{x}} \left(\frac{\partial}{\partial \theta} \log f(\vec{x}|\theta) \right) f(\vec{x}|\theta)$$

$$\therefore E_\theta \left(\frac{\partial}{\partial \theta} \log f(\vec{x}|\theta) \right) = 0$$

$$\text{ii)} \quad \text{Cor}(X, Y) \leq \text{Var}(X) \text{Var}(Y)$$

$$\Rightarrow \text{Note that: } \frac{\lambda}{\lambda\theta} E_\theta (W(\vec{x})) = E_\theta (W(\vec{x})) \frac{\partial}{\partial \theta} \log f(\vec{x}|\theta) \\ = \text{Cor}(W(\vec{x}), \frac{\partial}{\partial \theta} \log f(\vec{x}|\theta)). \text{ Since by ii) } \square.$$

Cir. If $X_k \sim f(x|\theta), 1 \leq k \leq n$, i.i.d. under the condition above. Then:

$$\text{Var}_\theta(W(\vec{x})) \geq \frac{\left(\frac{\lambda}{\lambda\theta} E_\theta (W(\vec{x})) \right)^2}{n E \left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2}$$

$$\begin{aligned} \underline{\text{Pf: Note}} \quad & E_{\theta} \left(\frac{\partial}{\partial \theta} \log f(x_i | \theta) \cdot \frac{\partial}{\partial \theta} \log f(x_j | \theta) \right), \quad i \neq j \\ & = E_{\theta} \left(\frac{\partial}{\partial \theta} \log f(x_i | \theta) \right) E_{\theta} \left(\frac{\partial}{\partial \theta} \log f(x_j | \theta) \right) = 0 \end{aligned}$$

Remark: i) Replace "ʃ" by "Σ", we attain the discrete form.

ii) $E_{\theta} \left(\left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right)^2 \right)$ is the Fisher information of sample in θ-space.

⇒ A computational Result:

If $f(x | \theta)$ satisfies:

$$\frac{1}{n\theta} E_{\theta} \left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right) = \int_x \frac{\partial}{\partial \theta} \left[\left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right) f(x | \theta) \right] dx$$

$$\text{Then } E_{\theta} \left(\left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right)^2 \right) = - E_{\theta} \left(\frac{\partial^2}{\partial \theta^2} \log f(x | \theta) \right)$$

$$\underline{\text{Pf: Note}} \quad \frac{1}{n\theta} E_{\theta} \left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right) = 0$$

$$= \int_x \left[\frac{\partial^2}{\partial \theta^2} \log f(x | \theta) + \left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right)^2 \right] f(x | \theta) dx.$$

Shortcoming:

i) The value of Cramér-Rao Lower Bound may not be attained by $\hat{\theta} \in C_2$.

ii) Some r.v. may not satisfies regular condition.

C.R. $X_k \sim f(x | \theta)$, i.i.d. $1 \leq k \leq n$, satisfies conditions

in Cramér-Rao Thm. $L(\theta | \vec{x}) = \prod f(x_i | \theta)$

$W(X_1, X_2, \dots, X_n) \in C_2^\theta$. Then $W(\vec{X})$ attains the Cramér-Rao Lower Bound $\Leftrightarrow \exists a(\theta)$. $a(\theta)[W(\vec{X}) - I(\theta)] = \frac{\partial}{\partial \theta} \log L(\theta | \vec{x})$

Pf: By Cauchy-Schwarz Inequality.

Remark: If $X_k \sim f(x|\theta, M)$, M is unknown, then $a(\theta, M)$ isn't function of θ . If M is known.

Then $a(\theta, M) = a(\theta)$. ✓

Approach two:

- Relate the sufficient statistic to unbiased estimator.

Thm (Rao-Blackwell)

$W \in C_2^\theta$, T is s.s. of θ . Define $\phi(T) = E(W|T)$.

Then $\phi(T) \in C_2^\theta$, and $\text{Var}_\theta(\phi(T)) \leq \text{Var}_\theta(W)$. $\# \theta$.

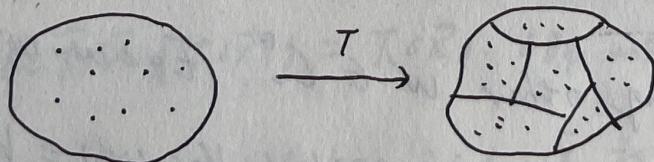
Pf: $E_\theta(W) = E_\theta(E(W|T)) = E_\theta(\phi(T)) = I(\theta)$

$$\text{Var}_\theta(W) = \text{Var}_\theta(E(W|T)) + E_\theta[\text{Var}(W|T)] \geq \text{Var}_\theta(\phi(T))$$

Moreover, $\phi(T)$ indep. of θ , so it's an estimator! (By S.S.)

Remark: An interpretation:

After the information comes in:



the partitioned elements have number-decreasing!

So Variance will decrease, definitely

\Rightarrow So we only need to consider the unbiased estimators condition on S.S!

But how do we know whether a UE is UMVUE?

Lemma. If w is UMVUE of $\theta(w)$. Then it's unique.

Pf: By contradiction: If w' is another UMVUE of $\theta(w)$.

Consider $w^* = \frac{1}{2}(w+w')$. $\text{Var}_\theta(w^*) \leq \text{Var}_\theta(w)$. By Cauchy.

$\therefore \text{Var}_\theta(w^*) = \text{Var}_\theta(w) \Rightarrow a(\theta)w^* + b(\theta) = w$. Determine a, b !

We obtain $w^* = w$. $w = w'$. Contradict!

\Rightarrow To see a UE is the best, we can check

if we can improve it:

Thm. $w \in C_2$. Then it is UMVUE $\Leftrightarrow \text{Cov}_\theta(w, u) = 0$

$\forall u$, satisfies $E_u(u) = 0$.

Pf: (\Rightarrow). If $\exists u$, $E_u(u) \neq 0$, $\text{Cov}_\theta(w, u) \neq 0$.

Let $w^* = u\alpha + w$. $\therefore \text{Var}_\theta(w^*) = \text{Var}_\theta(w) + 2\alpha \text{Cov}_\theta(u, w) + \alpha^2 \text{Var}_\theta(u)$

\therefore let $\alpha \in (0, -\frac{2\text{Cov}_\theta(u, w)}{\text{Var}_\theta(u)})$ or $(-\frac{2\text{Cov}_\theta(u, w)}{\text{Var}_\theta(u)}, 0)$

We have an improvement. When $\text{Var}_\theta(u) \neq 0$. (i.e. $u \neq 0$)

If $\text{Var}_\theta(u) = 0$, easy to see improvements exist!

(\Leftarrow) For any other $w' \in C_n$, $E_\theta(w-w') = 0$.

$\therefore \text{Cov}_\theta(w-w'; w) = 0$. Check $\text{Var}_\theta(w') \geq \text{Var}_\theta(w)$

Remark: i) W is called random noise, carrying no information.

ii) It's difficult to check the whole class of random noises. But it can be used to test W isn't UMVUE!

\Rightarrow Under some special condition, i.e. if for a family $f(x|\theta)$, it doesn't have random noises. Actually, this is the property of complete family.

However, recall that this property is called completeness!

Thm. T is a complete sufficient statistic for θ .

$\phi(T) \in C_2^\theta$. Then $\phi(T)$ is the unique UMVUE for $E(W|T)$.

Remark: i) To generate $\phi(T)$. Let $\phi(T) = E(W|T)$, $W \in C_2$.

ii) Interpretation: Completeness simplifies the estimation of θ , at most. Then $V_{\text{var}}(\phi(T))$ (error) can't be reduced any more.

iii) Sometimes $E(W|T)$ is difficult to calculate. By moments:
Then, we can find $E(T) = f(\theta)$. $\Rightarrow P(T) \in C_2^\theta$. Find P is poly.

Cor. $T(\bar{X})$ is complete sufficient statistic. $T(\bar{X}) \in C_2$.

For $\tilde{T}(\bar{X})$ is another statistic, s.t. $\tilde{T}(\bar{X}) \in C_2^\theta$.

Then $E(\tilde{T}(\bar{X}) | T(\bar{X})) = T(\bar{X})$. (By uniqueness)

e.g. $X_k \sim \text{Poisson}(\lambda)$, $k \in \mathbb{N}$, i.i.d. \bar{X} is complete sufficient statistic for λ . $E(S^2) = \lambda$, So we obtain:

$$E(S^2 | \bar{X}) = \bar{X}. \text{ An amazing result!}$$