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 $W(\vec{X})$. i.e. Any statistic is a point estimation.

(1) Method of finding estimations:

① Method of moments:

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

Denote $\mu_k = E(\hat{\mu}_k)$. $\hat{\mu}_k = \sum_1^n \frac{X_i^k}{n}$

Then
$$\mu_1 = \mu_1(\theta_1, \theta_2 \cdots \theta_k)$$
$$\mu_2 = \mu_2(\theta_1 \cdots \theta_k)$$
$$\vdots$$
$$\mu_k = \mu_k(\theta_1 \cdots \theta_k)$$

Replace μ_k by $\hat{\mu}_k$ (estimator)

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

Remark: A Technique: momenting matching:

e.g. (Satterthwaite Approximation)

$Y_i \sim \chi^2_{r_i}$. We want to find ν s.t. $\sum_1^k a_i Y_i$

will approximate $\frac{\chi^2_\nu}{\nu}$. But ν is unknown.

Note that: $\begin{cases} E(\sum a_i Y_i) = \sum a_i r_i = E(\frac{\chi_\nu^2}{\nu}) = 1 \\ E(\sum a_i Y_i)^2 = E(\frac{\chi_\nu^2}{\nu})^2 = \frac{2}{\nu} + 1 \end{cases}$

$$\Rightarrow \nu \approx \frac{2}{(\sum_1^k a_i Y_i)^2 - 1}$$, it's complicated and inefficient.

But to cancel the "1":

$$E((\sum a_i Y_i)^2) = Var(\sum a_i Y_i) + (E(\sum a_i Y_i))^2$$

$$= (E(\sum a_i Y_i))^2 \left(\frac{Var(\sum a_i Y_i)}{(E(\sum a_i Y_i))^2} + 1\right)$$

$$= \frac{Var(\sum a_i Y_i)}{(E(\sum a_i Y_i))^2} + 1 \cdot \frac{1}{\nu}$$ since $Var(D) = \sum_1^k \frac{a_i^2 E(Y_i)^2}{r_i}$

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

② Maximum Likelihood Estimators:

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

$$= \prod_1^n f(x_i | \theta_1 \cdots \theta_k)$$: X_i. 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 Sensity.

<u>For i)</u>: Let $\frac{\partial L(\theta|x)}{\partial \theta_i} = 0$. check $\frac{\partial^2}{\partial \theta_i^2} \ell \big|_{\hat\theta} < 0$

So that it's global maximum. ($\ell = \log L(\theta|x)$)

If θ must be an integer:

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

<u>Thm.</u> (Invariance property of MLEs)

If $\hat\theta$ is MLE of θ. Then for any func. $\tau(\cdot)$.
$\tau(\hat\theta)$ is MLE of $\tau(\theta)$.

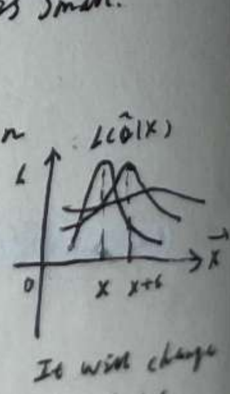
<u>Pf</u>: Define the induced Likelihood Func.

$$L^*(\eta|\vec{x}) = \sup_{\{\theta|\tau(\theta)=\eta\}} L(\theta|\vec{x}) . \quad L^*(\hat\eta|\vec{x}) = \sup_{\eta} \sup_{\{\theta|\tau(\theta)=\eta\}} L(\theta|\vec{x})$$

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

<u>For ii):</u>

We expect that: If base our calculation on $L(\theta|\vec{x}+\vec{\varepsilon})$. Then $\hat\theta_1$ should be closed to $\hat\theta$, for $\vec{\varepsilon}$ is small. ($\hat\theta_1$ is its MLE). However, this only happens when the likelihood Function is very flat in the neighbor 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.

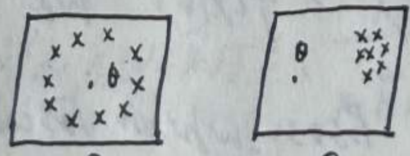
① **Mean Square Error:**

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

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

Actually, $E_\theta(W-\theta)^2 = Var_\theta(W) + (E_\theta(W)-\theta)^2$

$$\triangleq Var_\theta(W) + Bias_\theta^2(W)$$



① W_1 ② W_2

$\Rightarrow Var_\theta(W_1) > Var_\theta(W_2)$

But $Bias(W_1) < Bias(W_2)$

How to compare MSE?

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

$$\begin{cases} \text{Measure-Equivaria: } W(\vec{x}) \text{ estimates } \theta \Rightarrow \bar{g}(W(\vec{x})) \text{ estimates } \bar{g}(\theta) \\ \text{Formal-Invaria: } W(\vec{x}) \text{ estimates } \theta \Rightarrow W(g(\vec{x})) \text{ estimates } \bar{g}(\theta) \end{cases}$$

($\bar{g}: \mathbb{R} \to \mathbb{R}$. but $g: \mathcal{X} \to \mathcal{X}$. 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 minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ ✓

  If any other estimator W. $Var_\theta(W^*) \le Var_\theta(W)$. $\forall \theta$.

  where $E_\theta(W) = \tau(\theta)$. i.e. $W \in C_\tau = \{W \mid E_\theta(W) = \tau(\theta)\}$.

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

Approach one:

- If estimate $\tau(\theta)$. where $X \sim f(x|\theta)$. we can specify a lower bound $B(\theta)$ on the variance of any unbiased estimator of $\tau(\theta)$. Then we try to find W^* s.t. $Var_\theta(W^*) = B(\theta)$.

<u>Thm.</u> (Cramér - Rao Inequality)

$\vec{X} = (X_1 \cdots X_n) \sim f(\vec{x}|\theta)$. $W(\vec{X})$ an estimator satisfying regular condition: $\frac{d}{d\theta} E_\theta W(\vec{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta} [W(\vec{x}) f(\vec{x}|\theta)] d\vec{x}$.

And $Var_\theta(W(\vec{X})) < \infty$.

Then: $Var(W(\vec{X})) \geq \dfrac{(\frac{d}{d\theta} E_\theta(W(\vec{X})))^2}{E_\theta((\frac{\partial}{\partial\theta} \log f(\vec{X}|\theta))^2)}$

<u>Pf.</u> <u>Lemma.</u> i) An identity:

since $\int_{\mathcal{X}} f(x|\theta) dx = 1$. $\therefore \frac{\partial}{\partial\theta} \int_{\mathcal{X}} f(x|\theta) dx = 0$

$LHS = \int_{\mathcal{X}} \frac{\partial}{\partial\theta} f(x|\theta) dx = \int_{\mathcal{X}} (\frac{\partial}{\partial\theta} \log f(x|\theta)) f(x|\theta)$

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

ii) $Cov^2(X, Y) \leq Var(X) Var(Y)$

$\Rightarrow$ Note that: $\frac{d}{d\theta} E_\theta(W(X)) = E_\theta(W(X) \frac{\partial}{\partial\theta} \log f(x|\theta))$

$= Cov(W(X), \frac{\partial}{\partial\theta} \log f(x|\theta))$. Since by i) $\square$.

<u>Cor.</u> If $X_k \sim f(x|\theta)$. $1 \leq k \leq n$. i.i.d. under the condition above. Then:

$Var(W(\vec{X})) \geq \dfrac{(\frac{d}{d\theta} E_\theta(W(\vec{X})))^2}{n E((\frac{\partial}{\partial\theta} \log f(x|\theta))^2)}$

Pf: Note $E_\theta(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\frac{\partial}{\partial\theta}\log f(X_j|\theta))$. $i\neq j$

$$= E_\theta(\frac{\partial}{\partial\theta}\log f(X_i|\theta))\, E_\theta(\frac{\partial}{\partial\theta}\log f(X_j|\theta)) = 0$$

Remark: i) Replace "$\int$" by "Σ": we attain the discrete form.

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

$\Rightarrow$ A computational result:

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

$$\frac{d}{d\theta}E_\theta(\frac{\partial}{\partial\theta}\log f(x|\theta)) = \int_{\mathcal{X}}\frac{\partial}{\partial\theta}[(\frac{\partial}{\partial\theta}\log f(x|\theta))f(x|\theta)]\,dx$$

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

Pf: Note $\frac{d}{d\theta}E_\theta(\frac{\partial}{\partial\theta}\log f(x|\theta)) = 0$

$$= \int_{\mathcal{X}}[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) + (\frac{\partial}{\partial\theta}\log f(x|\theta))^2]\,f(x|\theta)\,dx.$$

Shortcoming:

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

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

Cor. $X_k \sim f(x|\theta)$. i.i.d. $1\le k\le n$. satisfies conditions in Cramér-Rao Thm. $L(\theta|\vec{x}) = \prod f(x_i|\theta)$

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

Pf: By Cauchy-Schwartz Inequality.

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

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 $Var_\theta(\phi(T)) \le Var_\theta(W)$. $\forall\theta$.

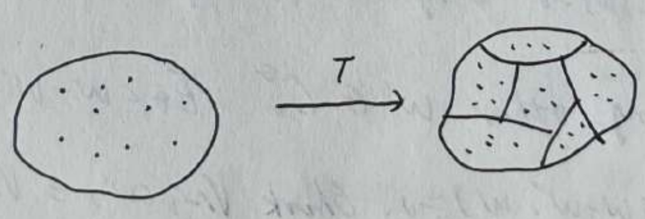
Pf: $E_\theta(W) = E_\theta(E(W|T)) = E_\theta(\phi(T)) = \tau(\theta)$

$Var_\theta(W) = Var_\theta(E(W|T)) + E_\theta(Var(W|T)) \ge Var_\theta(\phi(T))$

Moreover. $\phi(T)$ indept of θ. So it's an estimator! (By S.S)

Remark: An interpretion:

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?

<u>Lemma.</u> If W is UMVUE of $\tau(\theta)$. Then it's unique.

<u>Pf:</u> By contradiction: If W' is another UMVUE of $\tau(\theta)$.

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

$\therefore Var_\theta(W^*) = Var(W) \Rightarrow a(\theta)W^* + b(\theta) = W$. 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:

<u>Thm.</u> $W \in C_2^\theta$. Then it is UMVUE $\Leftrightarrow$ $Cov_\theta(W, U) = 0$

$\forall U$. satisfies $E_\theta(U) = 0$.

<u>Pf:</u> $(\Rightarrow)$. If $\exists U$. $E_\theta(U) = 0$. $Cov_\theta(W, U) \neq 0$.

Let $W^* = Ua + W$. $\therefore Var(W^*) = Var_\theta(W) + 2a\, Cov_\theta(U, W) + a^2 Var_\theta(U)$

$\therefore$ Let $a \in (0, -\frac{2 Cov_\theta(W, U)}{Var_\theta(U)})$ or $(-\frac{2 Cov_\theta(W, U)}{Var_\theta(U)}, 0)$

we have an improvement. when $Var_\theta(U) \neq 0$. (i.e. $U \neq 0$)

If $Var_\theta(U) = 0$. easy to see improvements exist!

$(\Leftarrow)$ For any other $W' \in C_2^\theta$. $E_\theta(W - W') = 0$.

$\therefore Cov(W - W', W) = 0$. Check $Var_\theta(W') \geq Var_\theta(W)$

Remark: i) U 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. $\phi(T) \in C_\tau^\theta$. Then $\phi(T)$ is the unique UMVUE for $\tau(\theta)$.

Remark: i) To generate $\phi(T)$. Let $\phi(T) = E(W|T)$, $W \in C_\tau^\theta$.

ii) Interpretation: completeness simplifies the information of θ at most. Then $Var_\theta(\phi(T))$ (error) can't be reduced any more.

iii) Sometimes $E(W|T)$ is difficult to calculate. By moments: Then we can find $E(T^k) = f_k(\theta)$. $\Rightarrow p(T) \in C_\tau^\theta$. Find p, is poly.

Cor. $T(\vec{x})$ is complete sufficient statistic. $T(\vec{x}) \in C_\tau^\theta$. For $\tilde{T}(\vec{x})$ is another statistic, st. $\tilde{T}(x) \in C_\tau^\theta$. Then $E(\tilde{T}(\vec{x}) | T(\vec{x})) = T(\vec{x})$. (By uniqueness)

e.g. $X_k \sim Poisson(\lambda)$, $1 \le k \le n$, i.i.d. $\bar{X}$ is complete sufficient statistic for λ. $E(S^2) = \lambda$. So we obtain:

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

Choice of Estimation Method.

$x_1, x_2 \cdots x_n$ (values) $\longrightarrow$ $\hat{\theta}_0$ (value)

$\uparrow$ realization

$X_1 \cdots X_n$ (r.v.'s) $\xrightarrow{\text{Func.}}$ $\hat{\theta}$ (r.v) $\xrightarrow{\text{Point Estimate}}$ θ (parameter)

Histogram

$\updownarrow$ $\uparrow$ Let $\theta = \hat{\theta}_0$ Simulation

$F_{\vec{X}}(\cdot|\theta)$

cdf: $F_{\hat{\theta}}(\cdot|\theta)$ of $\hat{\theta}$. $\longrightarrow$ cdf: $F_{\hat{\theta}}(\cdot|\hat{\theta}_0)$

θ is unknown

Normal cdf

$\Downarrow$

Standard Error $\sigma(\theta)$

$\Downarrow$

Estimated Standard Error: $\sigma(\hat{\theta}_0)$

① Exact Dist: The form of $F_{\hat{\theta}}(\cdot|\theta)$ is known.

② Asymptotical Method: when n is large: $F_{\hat{\theta}}(\cdot|\theta) \longrightarrow \Phi$

③ Simulation: The form of $F_{\hat{\theta}}(\cdot|\theta)$ is unknown. n is small.