Module – III

Parametric Point Estimation

$X_1, \ldots, X_n$: random sample from a population described by family $\{f_\theta(\cdot) : \theta \in \Theta\}$ of pdfs/pmfs, where form of $f_\theta(\cdot)$ is known for each $\theta \in \Theta$, but $\theta \in \Theta$ is unknown.

$P=$

Here knowledge of unknown $\theta \in \Theta$ yields knowledge of the entire population. Moreover θ itself may represent an important characteristic of the population (such as population mean, variance etc.) and there may be a direct interest in obtaining a point estimate of θ. Sometimes there may be interest in obtaining ~~a~~ point estimate of $\tau(\theta)$, for a given real or vector valued function of ~~θ~~ $\tau(\theta)$.

Goal: Based on a random sample $X_1, \ldots, X_n$ from the population, find a good point estimator of $\tau(\theta)$.

Definition A point estimator of $\tau(\theta)$ is a function $W(\underline{X})$ of the random sample $\underline{X} = (X_1, \ldots, X_n)$.

$\tau(\theta) = P_\theta(X > 1)$

Estimator ↔ Rule

Note: (i) An estimator $W(\underline{X})$ is a random variable whereas an estimate is an observed value of the estimator based on an observed sample.

(ii) An estimator $W(\underline{X})$ may not take values on Θ although it may be undesirable. It is done so in order to have a general discussion

Different Methods of Estimation

I. Method of Moments (Karl Pearson)

Suppose that $\underline{\theta} = (\theta_1, \dots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$.

Sometimes more than k equations will be required, e.g. $U(0,\theta)$

Equate first k sample moments to first k population moments and solve the resulting system of simultaneous equations.

Define

$$m_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j, \quad j = 1, 2, \dots, k$$

$$\mu_j'(\underline{\theta}) = E_{\underline{\theta}}(X_i^j), \quad j = 1, \dots, k.$$

Consider system of equations

$$\mu_j'(\underline{\theta}) = m_j, \quad j = 1, \dots, k$$

The solution $\tilde{\underline{\theta}} = (\tilde{\theta}_1, \dots, \tilde{\theta}_k)$ of above system of equations is the method of moment estimator (MME) of $\underline{\theta}$. $\tau(\tilde{\underline{\theta}})$ is called the MME of $\tau(\underline{\theta})$.

MMEs may not be the best estimators and may be improved upon.

Example 1

$\underline{\theta} = (\mu, \sigma)$

$\mu_1'(\underline{\theta}) = E_{\underline{\theta}}(X_1) = \mu$

$\mu_2'(\underline{\theta}) = E_{\underline{\theta}}(X_1^2) = \sigma^2 + \mu^2$

$m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

$m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

MME is $(\tilde{\mu}, \tilde{\sigma}^2)$

$\tilde{\mu} = \bar{X}$

$\tilde{\mu}^2 + \tilde{\sigma}^2 = \frac{1}{n}\sum X_i^2$

$\tilde{\sigma}^2 = \frac{1}{n}\sum X_i^2 - \bar{X}^2 = \frac{1}{n}\sum (X_i - \bar{X})^2$

$(\tilde{\mu}, \tilde{\sigma}^2) = (\bar{X}, S_n^2)$

MME of $\tau(\underline{\theta}) = \tilde{\mu} + \tilde{\sigma}$ is $\tilde{\tau} = \bar{X} + S_n$

Example 1 Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ where $\underline{\theta} = (\mu, \sigma)$ $\in \mathbb{R} \times (0, \infty) = \Theta$ is unknown. Find MME of $\underline{\theta}$. Is it based on minimal sufficient statistic?

Ex. $X_1, \dots, X_n$ iid $U(0,\theta)$

$\mu_1'(\theta) = E(X_1) = \frac{\theta}{2}$

$\frac{\tilde{\theta}}{2} = \bar{X}$

$\tilde{\theta} = 2\bar{X}$ not suff.

Example 2 Let $X_1, \dots, X_n$ be a random sample from $\text{Bin}(m, p)$ unknown. Find with both $m \in \mathbb{N}$ and $p \in (0,1)$ unknown. Find MME of $\underline{\theta} = (m, p) \in \mathbb{N} \times (0,1) = \Theta$ where $\mathbb{N} = \{1, 2, \dots\}$. Is it based on minimal sufficient statistic?

$\mu_1'(\underline{\theta}) = mp$; $\tilde{m}\tilde{p} = \bar{X}$

$\mu_2'(\underline{\theta}) = mp(1-p)$; $\tilde{m}\tilde{p}(1-\tilde{p}) + \tilde{m}^2\tilde{p}^2 = \frac{1}{n}\sum X_i^2$

$(\tilde{m}, \tilde{p})$ depends on $(\bar{X}, \frac{1}{n}\sum X_i^2)$

$f_{m,p}(\underline{x}) = \prod_{i=1}^{n} \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} I(x_i \in \{0, 1, \dots, m\})$

$= \frac{(m!)^n}{\prod_{i} x_i! \prod_{i} (m-x_i)!} p^{\sum x_i} (1-p)^{nm - \sum x_i} \prod_{i=1}^{n} I\{\dots\}$

minimal sufficient statistic $\left(\sum X_i, \prod_{i=1}^{n} X_i!(m-X_i)!\right)$ $\prod_{i=1}^{n} X_i!(m-X_i)!$

(B) Method of Maximum Likelihood Estimation

For a given (observed) sample point $\underline{x} = (x_1, \dots, x_n)$, define

$$L_{\underline{x}}(\underline{\theta}) = \prod_{i=1}^{n} f_{\underline{\theta}}(x_i) \qquad \underline{\theta} \in \Theta,$$

$P_\theta(X_1 = x_1, \dots, X_n = x_n)$

as a function of $\underline{\theta}$.

$L_{\underline{x}}(\underline{\theta})$: The probability that the observed sample point $\underline{x}$ (sample point) came from population represented by pdf/pmf $f_{\underline{\theta}}(\underline{x})$, $\underline{\theta} \in \Theta$

It makes sense to find $\hat{\underline{\theta}}$ that maximizes $L_{\underline{x}}(\underline{\theta})$, for a given sample point $\underline{x}$, as then the corresponding population (represented by pdf/pmf) is most likely to have yielded the observed sample $\underline{x}$.

Definition (I) For a given sample point $\underline{x}$, the function $L_{\underline{x}}(\underline{\theta})$, as a function of $\underline{\theta} \in \Theta$, is called the likelihood function.

(II) For each $\underline{x} \in \mathcal{X}$, let $\hat{\underline{\theta}} = \hat{\underline{\theta}}(\underline{x})$ be such that

$$L_{\underline{x}}(\hat{\underline{\theta}}) = \sup_{\underline{\theta} \in \Theta} L_{\underline{x}}(\underline{\theta}).$$

Then a maximum likelihood estimator (MLE) of the parameter $\underline{\theta} = (\theta_1, \dots, \theta_k)$, based on random sample $\underline{X}$, is $\hat{\underline{\theta}}(\underline{X})$.

$\theta \in (0,1)$
Ex $B(1, \theta)$ case
MLE may take value 0
MLE $\bar{X}$

Remark: (a) The range of MLE $\hat{\theta}(\underline{X})$ coincides with the range Θ of $\underline{\theta} = (\theta_1, \dots, \theta_k)$ with boundaries of Θ included.

(b) MLE is the parameter point (or the corresponding pdf/pmf) for which the observed sample is most likely.

(c) In general MLE possesses good optimality properties and hence a good point estimator.

$\boxed{3\tfrac{1}{2}}$

(d) Finding MLE requires maximization of likelihood function which is sometimes difficult and may require numerical optimization techniques. Ex $X_1, \ldots, X_n$ r.s. from Gam(α, θ), where $(\alpha, \theta) \in (0,\infty) \times (0,\infty) = \Theta$. Find the MLE of $(\hat{\alpha}, \hat{\theta})$

(e) MLE is sometimes sensitive to data; a slightly different data may produce a vastly different MLE making its use suspicious.

$L_x(\underline{\theta}) = \prod f_\theta(x_i) = \psi(\theta, T(x))\, h(\underline{x})$

(b) Since $L_{\underline{x}}(\underline{\theta})$ depends on sufficient (in fact minimal sufficient statistic, if it exists), the MLE is a function of every sufficient/minimal suff. stat.

Finding MLE:

Suppose that, for given $\underline{x} \in \mathcal{X}$, $L_{\underline{x}}(\underline{\theta})$ is a differentiable function of $\underline{\theta}$. Then the MLE $\hat{\underline{\theta}} = \hat{\underline{\theta}}(\underline{x})$ is either on the boundary of Θ or at a point where

$$\left[\frac{\partial}{\partial \theta_i} L_{\underline{x}}(\underline{\theta}) \right]_{\underline{\theta} = \hat{\underline{\theta}}(\underline{x})} = 0 \qquad i = 1, \ldots, k; \quad \cdots \text{(I)}$$

here $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta$.

Note that the points where (I) holds may be a local or global maxima or a saddle point. Solutions of (I) should be checked for global maxima.

Note that maximizing $L_{\underline{x}}(\underline{\theta})$ is equivalent to maximizing $l_{\underline{x}}(\underline{\theta}) = \ln L_{\underline{x}}(\underline{\theta})$. Sometimes it is convenient to consider maximization of $l_{\underline{x}}(\underline{\theta})$.

Example Let $X_1, \ldots, X_n$ be a random sample from a pdf/pmf $f_\theta(\cdot)$ where $\theta \in \Theta \subseteq \mathbb{R}$ is unknown. In each of the following case, find the MLE of θ.

(a) $X_1 \sim N(\theta, 1)$, $\Theta = \mathbb{R}$;

(b) $X_1 \sim N(\theta, 1)$, $\Theta = [0, \infty)$ $\rightarrow$ $l_{\underline{x}}(\underline{\theta}) = C(\underline{x}) + n\bar{x}\theta - \frac{n\theta^2}{2}$, $\theta \geq 0$

$\frac{\partial l_{\underline{x}}(\theta)}{\partial \theta} = n\bar{x} - n\theta = 0$ $\hat{\theta} = \bar{x}$

(Restricted MLE)

$\frac{\partial^2 l_{\underline{x}}(\theta)}{\partial \theta^2} = -n < 0$

$\hat{\theta}_{MLE} = \begin{cases} \bar{x}, & \bar{x} > 0 \\ 0 & \bar{x} \leq 0 \end{cases}$

(c) $X_1 \sim Bin(m, \theta)$ m is known and $\theta \in \Theta = (0, 1)$

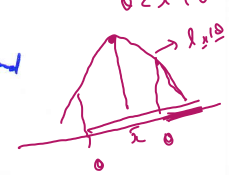
Aliter (↓)

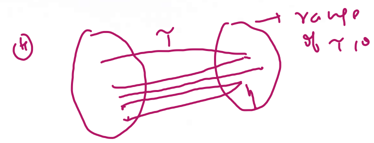
$l_{\underline{x}}(\theta) \uparrow$ 166

$\theta < \bar{x}$ ($\theta > \bar{x}$)

(d) $X_1 \sim Bin(m, \theta)$ m is known and $\theta \in \Theta = [0, 1]$

(e) $X_1 \sim Bin(m, p_0)$ where $p_0 \in (0, 1)$ is ~~un~~known and $m \in \{1\} = \Theta$ is unknown.





Invariance of MLE

Suppose that one is interested in estimating $\tau(\underline{\theta})$ for some function $\tau(\cdot)$.

Definition (Induced likelihood function). For a given sample observation $\underline{x}$, the likelihood function induced by $\eta = \tau(\underline{\theta})$ is defined by

$$L_{\underline{x}}^{*}(\eta) = \sup_{\{\underline{\theta} \in \Theta : \eta = \tau(\underline{\theta})\}} L_{\underline{x}}(\underline{\theta}), \quad \eta \in \Theta^{*} = \{\tau(\underline{\theta}) : \theta \in \Theta\}$$

$\hookrightarrow$ range of $\tau(\theta)$

The value $\hat{\eta} \equiv \hat{\eta}(\underline{x})$ that maximizes $L_{\underline{x}}^{*}(\eta)$ will be called the MLE of $\eta = \tau(\underline{\theta})$.

Note:

$$\sup_{\eta \in \Theta^{*}} L_{\underline{x}}^{*}(\eta) = \sup_{\eta} \sup_{\{\underline{\theta} \in \Theta : \eta = \tau(\underline{\theta})\}} = \sup_{\underline{\theta} \in \Theta} L_{\underline{x}}(\underline{\theta})$$

Theorem (Invariance Property of MLEs): If $\hat{\underline{\theta}} \equiv \hat{\underline{\theta}}(\underline{x})$ is the MLE of $\underline{\theta}$, then $\tau(\hat{\underline{\theta}})$ is the MLE of $\tau(\underline{\theta})$.

Proof Fix $\underline{x} \in \mathcal{X}$. Let $\hat{\eta} \equiv \hat{\eta}(\underline{x})$ maximizes $L_{\underline{x}}^{*}(\eta)$. We are required to show that

$$L_{\underline{x}}^{*}(\hat{\eta}) \leq L_{\underline{x}}^{*}(\tau(\hat{\theta})).$$

We have

$$L_{\underline{x}}^{*}(\hat{\eta}) = \sup_{\eta} L_{\underline{x}}^{*}(\eta)$$
$$= \sup_{\eta} \sup_{\{\theta : \eta = \tau(\theta)\}} L_{\underline{x}}(\underline{\theta})$$
$$= \sup_{\theta} L_{\underline{x}}(\underline{\theta}) = L_{\underline{x}}(\hat{\theta})$$

Also

$$L_{\underline{x}}^{*}(\hat{\eta}) = L_{\underline{x}}(\hat{\theta}) \leq \sup_{\{\theta : \tau(\theta) = \tau(\hat{\theta})\}} L_{\underline{x}}(\underline{\theta}) \quad \left(\text{Since } \hat{\theta} \in \{\theta \in \Theta : \tau(\theta) = \tau(\hat{\theta})\}\right)$$

$\tau(\hat{\theta}) = \tau(\hat{\theta})$

$$= L_{\underline{x}}^{*}(\tau(\hat{\theta}))$$

Hence $L_{\underline{x}}^{*}(\hat{\eta}) = L_{\underline{x}}(\hat{\theta}) \leq L_{\underline{x}}^{*}(\tau(\hat{\theta}))$. $\Rightarrow$ $\hat{\eta} \equiv \tau(\hat{\theta})$, $L_{\underline{x}}^{*}(\hat{\eta}) = L_{\underline{x}}^{*}(\tau(\hat{\theta}))$

Example (a) Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$, $\theta \in \Theta$. Then $\bar{X}^2$ is the MLE of θ^2. $\hat{\theta}_{MLE} = \bar{X}$ is MLE of θ^2 will be $\hat{\eta} = \hat{\theta}^2_{MLE} = \bar{X}^2$

(b) Let $X \sim Bin(n, \theta)$ where $n \in \mathbb{N}$ is known and $\theta \in \Theta$ is unknown. $\hat{\theta}_{MLE} = \frac{X}{n}$ Then the MLE of $\sqrt{\theta(1-\theta)} = s.d.(X_1)$ is

$$\sqrt{\frac{X}{n}\left(1-\frac{X}{n}\right)}$$

Result: Suppose $D \subseteq \mathbb{R}^n$ and $f: D \to \mathbb{R}$. Suppose $\frac{\partial f(\underline{x})}{\partial x_i}$, $i = 1, \ldots, n$ are continuous and have continuous partial derivatives on D. Let $\underline{x}_0$ be an interior point of D with

$$\left[\frac{\partial f(\underline{x})}{\partial x_i}\right]_{\underline{x} = \underline{x}_0} = 0, \quad i = 1, \ldots, n$$

Suppose

$$H = \left(\frac{\partial^2 f(\underline{x})}{\partial x_i \, \partial x_j}\right)_{\underline{x} = \underline{x}_0}$$

is the $n \times n$ Hessian matrix.

(a) If H is p.d. (or equivalently has all eigen values positive) then $\underline{x}_0$ is a point of local minima.

(b) If H is n.d. (or equivalently has all eigen values negative) then $\underline{x}_0$ is a point of local maxima.

(c) If H has both positive and negative eigen values then $\underline{x}_0$ is a saddle point. (For $n=2$, we get saddle point if $|H| < 0$)

Note: Let

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Then

(a) A is p.d. $\iff$ $a > 0$ and $ac - b^2 > 0$

(b) A is n.d. $\iff$ $a < 0$ and $ac - b^2 > 0$

~~(c) A has both positive and negative eigen values~~

Example Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, where $\underline{\theta} = (\mu, \sigma) \in \mathbb{R} \times (0, \infty) = \Theta$ is unknown. Show that $(\hat{\mu}, \hat{\sigma}^2) = \left(\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$ is the MLE of $\underline{\theta}$.

Example (MLE may not exist in strict sense) Let $X_1, \ldots, X_n$ be i.i.d. $Bin(1, \theta)$, where $\theta \in \Theta = (0, 1)$ is unknown. Here $T = \frac{\sum_{i=1}^{n} X_i}{n}$ is the MLE of θ. For $\sum_{i=1}^{n} X_i = 0$, $T = 0 \notin \Theta$ and for $\sum_{i=1}^{n} X_i = n$, $T = 1 \notin \Theta$.

Loss Functions for Evaluating Estimators

Any parametric estimation problem has following two basic components:

(i) Estimand: A real valued (or vector valued) function $\psi(\cdot)$ defined on parameter space Θ, whose value at θ is to be estimated.

(ii) A random variable (or random vector) $\underline{X}$ taking values on sample space $\mathcal{X}$ according to pdf/pmf f_θ, which is known to belong to family $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$, here functional form of $f_\theta(\cdot)$ is known but $\theta \in \Theta$ is unknown.

Goal: To find an estimator $\delta : \mathcal{X} \to \psi(\Theta)$, where $\psi(\Theta) = \{\psi(\theta) : \theta \in \Theta\}$

It is expected that the estimator $\delta(\underline{X})$ will be close to $\psi(\theta)$. For this it is necessary to specify a measure of closeness of an estimator to $\psi(\theta)$.

Suppose that the consequences of estimating $\psi(\theta)$ by a value d are measured by $L(\theta, d)$, $\theta \in \Theta$, $d \in \psi(\Theta)$. The function $L(\theta, d)$ is called the loss function. We shall assume that

$$L(\theta, d) \geq 0, \quad \forall\ \theta \in \Theta,\ d \in \psi(\Theta)$$

and

$$L(\theta, \psi(\theta)) = 0 \quad \forall\ \theta \in \Theta$$

For an estimator $\delta(x)$, $L(\theta, \delta(x))$ is a random variable and thus it makes sense to talk about average loss of an estimator $\delta(x)$ as

$$R(\theta, \delta) = E_\theta(L(\theta, \delta(x))), \quad \theta \in \Theta$$

We call $R(\theta, \delta)$, $\theta \in \Theta$, the risk function of estimator δ.

Randomized Estimators

We know that if T is a sufficient statistic then from the knowledge of T alone it is possible to construct a quantity $\underline{Y} \stackrel{d}{=} \underline{X}$, $\forall \theta \in \Theta$. Thus $\delta(\underline{X}) \stackrel{d}{=} \delta(\underline{Y})$, $\forall \theta$. Here the construction of $\underline{Y}$ depends on random mechanism. Thus $\delta(\underline{Y})$ not only depends on T but also a random mechanism. It is not an usual estimator defined ~~as~~ before, but a randomized estimator.

Definition Let X be the basic observable. A randomized estimator of $\psi(\theta)$ is a rule which assigns to each observed values x of X a random variable Y_x (or equivalently a known prob. dist p_x) with a known probability distribution. When $X = x$ is observed an observation of Y_x (from p_x) will be taken and will constitute the estimate of $\psi(\theta)$.

The risk of a randomized estimator $\delta(X)$ is given by

$$R(\theta, \delta) = E_\theta(r(\theta, \delta(X))),$$

where

$$r(\theta, \delta(x)) = E(L(\theta, Y_x)), \quad x \in \mathcal{X}, \ \theta \in \Theta.$$

Note: (1) If, for each $x \in \mathcal{X}$, Y_x is a degenerate r.v. (i.e. $P(Y_x = \delta(x)) = 1$, for some $\delta(\cdot)$) then the corresponding estimator is called a non-randomized estimator.

(ii) A non-randomized estimator gives the estimate of $\psi(\theta)$ immediately after observing $X=x$. A randomized estimator, after observing $X=x$, conducts a random experiment (of observing a r.v. Y_x from a known prob. dist p_x) and reports the outcome of this experiment as an estimator.

Theorem Let $T \equiv T(\underline{X})$ be a sufficient statistic and let $\delta(\underline{X})$ be an estimator (possibly randomized) of ~~estimator~~ estimand $\psi(\theta)$. Then there exists an estimator $\delta^*(T)$ (possibly randomized) based on sufficient statistic $T \equiv T(\underline{X})$ such that

$$R(\theta, \delta^*) = R(\theta, \delta) \quad \forall\ \theta \in \Theta.$$

Proof. Let $p_x(a)$ be the pdf/pmf corresponding to $\delta(x)$, $x \in \mathcal{X}$. Let $f_{X|T=t}(x|\theta)$ denote the pdf/pmf of X given $T \equiv T(X) = t$, $\theta \in \Theta$. For observed value $T(\underline{X}) = t$ let $p_t^*(a)$ be the ~~prob~~ pdf/pmf such that

$$p_t^*(a) = \int p_x(a)\, f_{X|T=t}(dx|\theta)$$

and let the corresponding r.v. be denoted by Z_t.

For observed value $\underline{X} = \underline{x}$, let $\delta^*(\underline{x})$ be the randomized estimator corresponding to $Z_{T(\underline{x})}$. Then

$$r(\theta, \delta^*(\underline{x})) = \int L(\theta, a)\, p^*_{T(x)}(da)$$

$$\begin{aligned}
R(\theta, \delta^*) &= E_\theta\left(r(\theta, \delta^*(\underline{X}))\right) \\
&= E_\theta\left(\int L(\theta, a)\, p^*_{T(X)}(da)\right) \\
&= \iint L(\theta, a)\, p_t^*(da)\, g_\theta(dt),
\end{aligned}$$

where g_θ denotes the pdf/pmf of $T(\underline{X})$.

Thus

$$\begin{aligned}
R(\theta, \delta^*) &= \iiint L(\theta, a)\, p_x(da)\, f_{X|T=t}(dx|\theta)\, g_\theta(dt) \\
&= \iint L(\theta, a)\, p_x(da)\, f_\theta(dx) = R(\theta, \delta) \quad \forall\ \theta \in \Theta
\end{aligned}$$

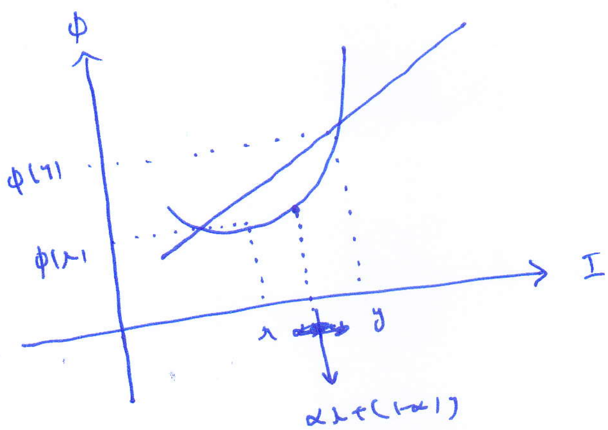
In most estimation problems the loss function is quite commonly a convex function

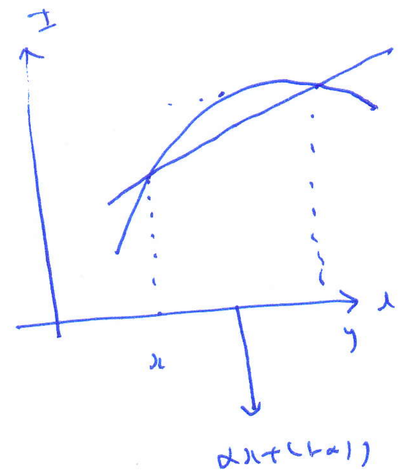
Definition: Let $\phi : I \to \mathbb{R}$ be a given function defined on an interval $I \subseteq \mathbb{R}$. The function ϕ is said to be convex (concave) on I if for every $x, y \in I$ and $\alpha \in (0, 1)$

$$\phi(\alpha x + (1-\alpha) y) \underset{(\ge)}{\le} \alpha\, \phi(x) + (1-\alpha)\, \phi(y). \quad \text{(concave)}$$

The function ϕ is said to be strictly convex (concave) if strict inequality holds above for all $x, y \in I$ and $\alpha \in (0, 1)$.

Note: ϕ is concave (convex) $\iff$ $-\phi$ is convex (concave)





Convex Function

Results: (a) If ϕ is convex on I then for every $x_1, x_2 \in I$ and $x \in (x_1, x_2)$

$$\phi(x) \le \frac{x_2 - x}{x_2 - x_1}\, \phi(x_1) + \frac{x - x_1}{x_2 - x_1}\, \phi(x_2)$$

(b) If ϕ is differentiable on I, then

ϕ is convex on I $\iff$ $\phi'(\cdot) \uparrow$ on I

(c) If ϕ is twice differentiable on I then

ϕ is convex on I $\iff$ $\phi''(x) \ge 0 \quad \forall x \in I$

(d) If ϕ is convex on I and if x_0 is an interior point of I then $\exists$ a real number $m \equiv m(x_0)$ such that

$$\phi(x) \geq \phi(x_0) + m(x - x_0) \quad \forall x \in I.$$

If ϕ is strictly convex, the inequality is strict for $x \in I$, $x \neq x_0$

(e) Every convex function is continuous.

Theorem (Jensen's Inequality). If ϕ is a convex function defined over an open interval I and X is a r.v. with $P(X \in I) = 1$ and finite expectation, then

$$E(\phi(X)) \geq \phi(E(X))$$

If ϕ is strictly convex the inequality is strict unless $P(X = c) = 1$, for some constant c.

<u>Examples</u>: The following functions are convex:

$\phi_1(x) = x^2$, $x \in \mathbb{R}$, $\phi_2(x) = |x - a|^p$, $p \geq 1$, $a \in \mathbb{R}$, $x \in \mathbb{R}$, $\phi_3(x) = e^x$, $x \in \mathbb{R}$, $\phi_4(x) = -\ln x$, $x \in (0, a)$, $\phi_5(x) = \frac{1}{x}$, $x \in (0, a)$.

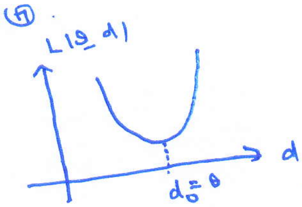
<u>Some inequalities</u>

(i) $$E(e^X) \geq e^{E(X)}$$

(ii) $$E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}$$

(iii) $$E(\ln X) < \ln E(X)$$

<u>Note</u>: It is reasonable to assume that any loss function $L(\theta, d)$ is convex in d for every $\theta \in \Theta$.



Some Convex Loss Functions

(I) Absolute Error Loss Function

$$L(\theta, d) = |d - \psi(\theta)|, \quad d \in \Theta^* = \psi(\Theta), \ \theta \in \Theta$$

(Symmetric Loss)

(II) Squared Error Loss Function

$$L(\theta, d) = (d - \psi(\theta))^2, \quad d \in \Theta^* = \psi(\Theta), \ \theta \in \Theta$$

(Symmetric Loss)

Symmetric loss functions assign equal weights to overestimation and underestimation of the same magnitude.

(III) Non-Symmetric Convex Loss Function

$$L(\theta, d) = e^{a(d - \psi(\theta))} - a(d - \psi(\theta)) - 1, \quad d \in \Theta^*, \ \theta \in \Theta,$$

where $a \neq 0$.

(IV) $$L(\theta, d) = |d - \psi(\theta)|^p, \quad d \in \Theta^*, \ \theta \in \Theta;$$

here $p > 0$ is a fixed number.

(V) For estimating $\underline{\psi}(\theta) = (\psi_1(\theta), \ldots, \psi_k(\theta))$

$$L(\underline{\theta}, \underline{d}) = \sum_{i=1}^{k} |d_i - \psi_i(\theta)|^p, \quad \underline{d} \in \Theta^*, \ \underline{\theta} \in \Theta,$$

here $p > 0$ is a fixed number.

Theorem Suppose that the loss function $L(\underline{\theta}, d)$ is a convex function of $d \in \mathbb{R}$, for every $\underline{\theta} \in \Theta \subseteq \mathbb{R}^k$. Let δ be a randomized estimator of the real-valued estimand $\psi(\underline{\theta})$. Then there exists a non-randomized estimator δ^* of $\psi(\underline{\theta})$ such that

$$R(\underline{\theta}, \delta^*) \leq R(\underline{\theta}, \delta), \quad \forall\, \underline{\theta} \in \Theta,$$

with strict inequality above if the loss function is strictly convex.

Proof Let Y_x be the r.v. corresponding to $\delta(x)$. Define

$$\delta^*(x) = E(Y_x), \quad x \in \mathcal{X}.$$

Clearly $\delta^*(x)$ is non-randomized. Also

$$\begin{aligned} M(\underline{\theta}, \delta(x)) &= E(L(\underline{\theta}, Y_x)) \\ &\geq L(\underline{\theta}, E(Y_x)) \\ &= L(\underline{\theta}, \delta^*(x)), \quad \forall\, \theta \in \Theta \end{aligned}$$

Thus

$$\begin{aligned} R(\underline{\theta}, \delta) &= E_\theta(M(\underline{\theta}, \delta(X))) \\ &\geq E_\theta(L(\underline{\theta}, \delta^*(X))) \\ &= R(\underline{\theta}, \delta^*), \quad \forall\, \theta \in \Theta \end{aligned}$$

Remark: In an estimation problem with convex loss function, given any randomized estimator $\delta(x)$, there exists a non-randomized estimator $\delta^*(\underline{X})$ that is as good as the randomized estimator $\delta(\underline{X})$.

Theorem (Rao–Blackwell). Let $T \equiv T(\underline{x})$ be a sufficient statistic for $\mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} \in \Theta\}$, where $\Theta \subseteq \mathbb{R}^k$. For estimating the real-valued estimand $\psi(\theta)$, suppose that the loss function $L(\theta, d)$ is strictly convex in d, for every $\theta \in \Theta$. Then given any non-randomized estimator $\delta^*(\underline{x})$ of $\psi(\theta)$ there exists a non-randomized estimator $\eta(T)$, based on the sufficient statistic $T \equiv T(\underline{x})$, such that

$$R(\underline{\theta}, \eta) \leq R(\underline{\theta}, \delta^*) \quad \forall\ \theta \in \Theta,$$

with strict inequality above if the loss function is strictly convex.

Proof. Define

$$\eta(T) = E_\theta(\delta^*(x) \mid T).$$

Clearly $\eta(\cdot)$ is $\eta(T)$ does not depend on θ (by virtue of sufficiency of T) and non-randomized. Moreover

$$R(\theta, \delta^*) = E_\theta(L(\underline{\theta}, \delta^*(x)))$$

$$= E_\theta(E_\theta(L(\theta, \delta^*(x)) \mid T))$$

But, by Jensen's inequality,

$$E_\theta(L(\theta, \delta^*(x)) \mid T) \geq L(\theta, E_\theta(\delta^*(x) \mid T))$$

$$= L(\theta, \eta(T))$$

Thus

$$R(\theta, \delta^*) \geq E_\theta(L(\theta, \eta(T))) = R(\theta, \eta).$$

Remark: In an estimation problem suppose that the loss function is strictly convex. Then, given any estimator (possibly randomized) $\delta(\underline{x})$, there exists a non-randomized estimator $\delta^*(T)$ based on sufficient statistic (or minimal sufficient statistic, if it exists) such that

$$R(\theta, \delta^*) \leq R(\theta, \delta) \quad \forall\ \theta \in \Theta,$$

With strict inequality above if the loss function is strictly convex.

Exercises

(1) Let $X_1, \ldots, X_n$ be i.i.d. $U(0, \theta)$ where $\theta \in \Theta = (0, \infty)$.

(a) For estimating $\psi(\theta) = \theta^2$ ~~quadratic~~ under squared error loss function $L(\theta, d) = (d-\theta)^2$, $d \in (0, \infty)$, $\theta \in (0, \infty)$ find an estimator better than $\delta_1(\underline{X}) = \frac{3}{n} \sum_{i=1}^{n} X_i^2$.

(b) For estimating $\psi(\theta) = \theta$ under absolute error loss function $L(\theta, d) = |d-\theta|$, $d \in (0, \infty)$, $\theta \in (0, \infty)$, find an estimator better than $\delta_2(\underline{X}) = \frac{2}{n} \sum_{i=1}^{n} X_i$.

(2) Let $X_1, \ldots, X_n$ be a random sample from Poisson(θ) where $\theta \in \Theta = (0, \infty)$. For estimating $\psi(\theta) = \theta$ under the absolute error loss function $L(\theta, d) = |d-\theta|$, $\theta, d \in (0, \infty)$ find an estimator better than $\delta_3(\underline{X}) = \frac{1}{n-1} \sum_{i=1}^{n-1} X_i$.

(3) Let $X \sim N(\theta, 1)$ and consider estimation of $\psi(\theta) = \theta$ under squared error loss function $L(\theta, d) = (d-\theta)^2$, $d, \theta \in \Theta = \mathbb{R}$. Let $\delta_0(x)$ be a randomized estimator of θ with corresponding randomizing r.v. as $Y_x \sim N(x, 1)$, $x \in \mathbb{R}$. Calculate the risk of δ_0 and find a non-randomized decision rule dominating δ_0.

Definition An estimator δ is said to be inadmissible if there exists another estimator δ^* which dominates it, i.e.

$$R(\theta, \delta^*) \le R(\theta, \delta) \quad \forall\, \theta \in \Theta$$

with strict inequality for some $\theta \in \Theta$.

An estimator which is not admissible is said to be admissible, i.e. an estimator δ is said to be admissible if for any estimator δ^* with

$$R(\theta, \delta^*) \le R(\theta, \delta) \quad \forall\, \theta \in \Theta$$

we have $R(\theta, \delta^*) = R(\theta, \delta) \quad \forall\, \theta \in \Theta$

Assignment Problems

(1) Show that every function of a complete statistic is complete.

(2) Show that the family $\mathcal{P}_1 = \{N(\theta, 1) : \theta \in \mathbb{R}\}$ is complete whereas the family $\mathcal{P}_2 = \{U(\theta, \theta+1) : \theta \in \mathbb{R}\}$ is not complete.

(3) Let $X_1, \ldots, X_n$ be i.i.d r.v.s with X_1 having pdf/pmf $f_\theta(\cdot)$, $\theta \in \Theta$. In each of the cases examine whether or not complete-sufficient statistic exist. If it exists, find it.

(a) $X_1 \sim Bin(n, \theta)$, n is known positive integer and $\Theta = (0, 1)$

(b) $X_1 \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\Theta = \mathbb{R}$

(c) $f_\theta(x) = e^{-(x-\theta)}$ $x > \theta \in \Theta = \mathbb{R}$; (d) $f_\theta(x) = \frac{2(\theta - x)}{\theta^2}$, $0 < x < \theta$.

(e) $f_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$, $x \in \mathbb{R}$, $\theta \in \Theta = \mathbb{R}$

(4) Show that if $\mathcal{P}$ is complete and $\mathcal{P}_0 \subseteq \mathcal{P} \subseteq \mathcal{P}_1$, then $\mathcal{P}_0$ and $\mathcal{P}_1$ may not be complete.

(5) Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distributions, respectively, where $\underline{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \in \Theta = \mathbb{R}^2 \times (0, \infty)$.

(a) Find a complete and sufficient statistic.

(b) If $\sigma_1^2 = \sigma_2^2$, find a complete and sufficient statistic.

(6) Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with pdf

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}}, \quad x > \mu,$$

where $\underline{\theta} = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$. Let $T_1 = X_{(1)}$ and $T_2 = \sum_{i=1}^{n} (X_i - X_{(1)})$. Show that $\underline{T} = (T_1, T_2)$ is a complete and sufficient statistic. Also show that T_1 and T_2 are statistically independent.

(7) Let $X_1, \ldots, X_n$ be iid $U(\theta_1, \theta_2)$ r.v.s where $\underline{\theta} = (\theta_1, \theta_2) \in \Theta = \{(x, y) \in \mathbb{R}^2 : -\infty < x < y < \infty\}$. Show that $T(\underline{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient. Also show that

$$\underline{Z} = \left(\frac{X_{(2)} - X_{(1)}}{X_{(n)} - X_{(1)}}, \ldots, \frac{X_{(n-1)} - X_{(1)}}{X_{(n)} - X_{(1)}} \right) \text{ and } T = X_{(1)} + X_{(n)}$$

are independent.

(8) Let X be a r.v. with p.m.f.

$$f_\theta(x) = \begin{cases} \frac{1}{4}, & x = 1, 2 \\ \frac{1}{4} + \frac{\theta}{4}, & x = 3 \\ \frac{1}{4} - \frac{\theta}{4}, & x = 4 \\ 0 & \text{o.w} \end{cases},$$

where $\theta \in \Theta = (0, 1)$. Find a minimal sufficient statistic for θ.

(9) Prove that a necessary and sufficient condition for a statistic T to be sufficient is that for any fixed θ and θ_0 in Θ, the ratio $\frac{f_\theta(x)}{f_{\theta_0}(x)}$ is a function of $T(\underline{x})$.

(10) Let $\mathcal{P} = \{f_0, f_1, \ldots, f_k\}$, where pdfs / pmfs f_i's have the same support. Prove that the statistic $T(\underline{x}) = \left(\frac{f_1(x)}{f_0(x)}, \ldots, \frac{f_k(x)}{f_0(x)} \right)$ is minimal sufficient.

(11) Let $\mathcal{P}$ be a family of distributions with common support and $\mathcal{P}_0 \subseteq \mathcal{P}$. If T is minimal sufficient for $\mathcal{P}_0$ and sufficient for $\mathcal{P}$, prove that T is minimal sufficient for $\mathcal{P}$.

(12) Let $X \sim Bin(n, \theta)$, $\theta \in (0, 1)$. For estimating θ under SEL function, find a non-randomized estimator dominating the randomized estimator $\delta(x)$ corresponding to r.v. Y_x with $Pr(Y_x = \frac{x}{n}) = Pr(Y_x = \frac{1}{2}) = \frac{1}{2}$. Also

$\boxed{16/3}$

calculate risk functions?

(13) Let $\mathcal{P} = \{f_0, f_1, \ldots, f_k\}$ and for any x let $S(x) = \{(i, j): f_i(x) + f_j(x) > 0\}$. Prove that the statistic $T(x) = \left\{ \frac{f_j(x)}{f_i(x)} : i < j,\ (i, j) \in S(x) \right\}$ is minimal sufficient. Here $\frac{f_j(x)}{f_i(x)} = \infty$ if $f_i(x) = 0$ and $f_j(x) > 0$.

(14) Let $\mathcal{P} = \{f_0, f_1, f_2\}$, where $f_0(x) = 1$, $-1 < x < 0$, $f_1(x) = 1$, $0 < x < 1$ and $f_2(x) = 2x$, $0 < x < 1$. Show that $\left(\frac{f_1(x)}{f_0(x)}, \frac{f_2(x)}{f_0(x)} \right)$ is not minimal sufficient. Is it sufficient? Find a minimal sufficient statistic.

(15) In an estimation problem suppose that the loss function is strictly convex.

(a) Show that any admissible estimator must be non-randomized.

(b) Suppose that δ is an admissible estimator of $\psi(\theta)$ and δ^* is any other estimator with $R(\theta, \delta^*) = R(\theta, \delta)$, $\forall\, \theta \in \Theta$. Show that $\delta^* = \delta$ w.p. 1.