Efficiency and Cramer-Rao inequality

Examples

8.7 of Rice

07/01/2021

Berkeley
UNIVERSITY OF CALIFORNIA

In the previous lecture,

WHICH ESTIMATOR OF $g(\theta)$ IS BEST?

- Asymptotic normality is stronger than consistency.
    - Pre-determine the sample size to ensure accuracy.
- Multi-parameter model and Fisher information matrix:
    - $N\left(\mu, \sigma^2\right)$
    - $\mathrm{Gamma}(\alpha, \beta)$
- Bootstrap confidence intervals using MLE.
    - Bonferroni correction.
- Asymptotic unbiasedness is easy to achieve.
    - $\mathrm{Var}\left(\hat{\theta}_n\right) \searrow$ Efficiency $\nearrow$
    - CR lower bound - the best possible estimator variance.
    - Under smooth assumptions, MLE always achieve the CR bound asymptotically.

Cramer-Rao inequality

Theorem E. Suppose the population has a density $f(x \mid \theta)$. Under the i.i.d and a few other assumptions, let $\hat{\delta}_n = g(X_1, \dots, X_n)$ be any estimator. Define $\psi(\theta) = E_\theta\left(\hat{\delta}_n\right)$. Then

$$\mathrm{Var}_\theta\left(\hat{\delta}_n\right) \geq \frac{[\psi'(\theta)]^2}{nI(\theta)},$$

where the lower bound is attained if and only if $\frac{\partial}{\partial\theta} l(\theta) = a(\theta)\left[\hat{\delta}_n - \psi(\theta)\right]$.

Candidate unbiased estimators for θ: $\hat{\theta}_n, \breve{\theta}_n, \mathring{\theta}_n, \cdots$

$$E_\theta\left(\hat{\theta}_n\right) = \theta, \text{ and CR bound} = \frac{1}{nI(\theta)}.$$

Candidate unbiased estimators for $g(\theta)$: $\hat{\delta}_n, \breve{\delta}_n, \mathring{\delta}_n, \cdots$

$$E_\theta\left(\hat{\delta}_n\right) = g(\theta), \text{ and CR bound} = \frac{[g'(\theta)]^2}{nI(\theta)}.$$

Cramer-Rao inequality

$EX = \lambda$, $\operatorname{var} X = \lambda$

Example 6. Let $X_1, \ldots, X_n$ be i.i.d $\text{Poisson}(\lambda)$. Consider all unbiased estimators of λ.

$$I(\lambda) = \frac{1}{\lambda}, \text{ and CR\_bound} = \frac{1}{nI(\lambda)} = \frac{\lambda}{n}.$$

$E(\bar{X}_n) = \lambda$

$\operatorname{var}(\bar{X}_n) = \frac{\operatorname{var}(X)}{n} = \frac{\lambda}{n}$.

Example 7. Let $X_1, \ldots, X_n$ be i.i.d $\text{Bernoulli}(p)$. Consider all unbiased estimators of p.

$f(x|p) = p^x (1-p)^{1-x}$

$\log f = x \log p + (1-x) \log(1-p)$

$\frac{\partial \log f}{\partial p} = \frac{x}{p} + \left(-\frac{1-x}{1-p}\right)$

$\frac{\partial^2 \log f}{\partial p^2} = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$

$I(p) = -E\left(\frac{\partial^2 \log f}{\partial p^2}\right) = \frac{EX}{p^2} + \frac{E(1-X)}{(1-p)^2}$

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

$= \frac{1}{p(1-p)}$.

Thus, CR_bound $= \frac{1}{nI(p)} = \frac{p(1-p)}{n}$.

$E\bar{X}_n = p$, $\operatorname{var}(\bar{X}_n) = \frac{\operatorname{var}(X)}{n} = \frac{p(1-p)}{n}$.

Cramer-Rao inequality

$E(\hat{\sigma}_n^2) = \tau(\sigma) = \sigma^2$, $\boxed{\frac{(n-1)\hat{\sigma}_n^2}{\sigma^2}} \sim \chi^2_{n-1}$

$var(\chi^2_{n-1}) = 2(n-1)$

Example 4 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$. Consider all unbiased estimators of μ and σ^2.

$$I(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

$$CR\_bound(\mu) = \frac{1}{n I(\mu)} = \frac{\sigma^2}{n}.$$

$$CR\_bound(\sigma^2) = \frac{[\tau'(\sigma)]^2}{n I(\sigma)} = \frac{[2\sigma]^2}{n\,{}^2/\sigma^2} = \underbrace{\frac{2\sigma^4}{n}}.$$

$$\frac{1}{n-1} > \frac{1}{n}$$

$E(\bar{X}_n) = \mu$

$$var(\bar{X}_n) = \frac{var(X)}{n} = \frac{\sigma^2}{n}.$$

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

$E(\hat{\sigma}_n^2) = \sigma^2.$

$$var(\hat{\sigma}_n^2) = var(\chi^2_{n-1}) \left(\frac{\sigma^2}{n-1}\right)^2$$

$$= \frac{\cancel{2(n-1)}\,\sigma^4}{(n-1)^{\cancel{2}}}$$

$$= \underbrace{\frac{2\sigma^4}{n-1}}.$$

$var(\hat{\sigma}_n^2) > CR\_bound(\sigma^2)$

CR-bound may not be attained even with UMVUE

Uniformly minimum-variance unbiased estimator (UMVUE) has lower variance than any other unbiased estimator.

Sufficiency

8.8 of Rice

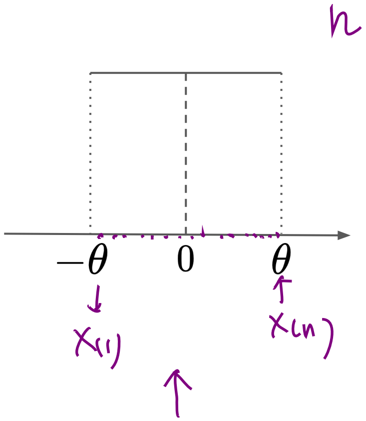
07/01/2021

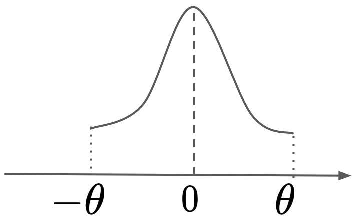
The assumptions do not always hold

Lec 4 Example 2. Let $X_1, \ldots, X_n$ be i.i.d $U(-\theta, \theta)$. Find the MLE for θ .

Lec 4 Example 3. Let $X_1, \ldots, X_n$ be i.i.d from a population with density

$$f(x \mid \theta) = \begin{cases} \frac{3x^2}{2\theta^3}, & \text{if } -\theta \leq x \leq \theta, \\ 0, & \text{otherwise.} \end{cases}$$





$$\hat{\theta}_{MLE} = \max\{-X_{(1)}, X_{(n)}\}$$

It is sufficient to only use $X_{(1)}$ and $X_{(n)}$ to infer θ!

Sufficient statistic

mathbf → bond font

For notational simplicity, denote $\mathbf{X}_n = (X_1, \ldots, X_n)$.

Definition. A statistic $T(\mathbf{X}_n)$ is *sufficient* for population parameter θ if $\mathbf{X}_n \mid T(\mathbf{X}_n)$ is independent of θ.

The joint density of $\mathbf{X}_n$		*The density of* $T(\mathbf{X}_n)$		*Conditional density of* $\mathbf{X}_n \mid T(\mathbf{X}_n)$
All *information contained in this sample*	—	*Information contained in this statistic*	=	*Remaining information*

θ

How to calculate conditional density?

joint density

$$h(y \mid X = x) = f(x, y)/g(x)$$

marginal density of X

$$\left\{\begin{aligned} &F(x,y) = P(X \le x, Y \le y) \\ &\frac{\partial^2 F}{\partial x \partial y} = f(x,y) \\ &g(x) = \int f(\cancel{x},y)\,dy \end{aligned}\right.$$

Sufficient statistic - *Directly use the definition*

Example 7 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $\mathrm{Bernoulli}(p)$. Is $\sum_{i=1}^{n} X_i$ sufficient for p?

Binomial $(2, p)$

Solution. $(X_1, \cdots X_n) \mid \sum_{i=1}^{n} X_i$

$$P(X_1+X_2=t) = \begin{cases} (1-p)^2 = P(X_1=0, X_2=0), & t=0 \\ 2p(1-p), & t=1 \\ p^2, & t=2 \end{cases}$$

$n=2$, $X_1 = \begin{cases} 0, & 1-p \\ 1, & p \end{cases}$ $X_2 = \begin{cases} 0, & 1-p \\ 1, & p \end{cases}$, indep

$$P(X_1 = x_1, X_2 = x_2 \mid X_1 + X_2 = t)$$

$x_1 \to 0, 1$, $x_2 \to 0, 1$, $t \to 0, 1, 2$

$$= \frac{P(X_1 = x_1, X_2 = x_2, X_1 + X_2 = t)}{P(X_1 + X_2 = t)} = \begin{cases} \dfrac{P(X_1 = x_1, X_2 = x_2, X_1 + X_2 = 0)}{(1-p)^2} = \begin{cases} 0, & \text{if } x_i = 1, i = 1, 2 \quad t=0 \\ \dfrac{P(X_1=0, X_2=0)}{(1-p)^2} = 1, & t=0 \end{cases} \\ \dfrac{P(X_1 = x_1, X_2 = x_2)}{2p(1-p)} = \dfrac{p(1-p)}{2p(1-p)} = \dfrac{1}{2}. \\ \cdots \quad t = 2 \end{cases}$$

$n \to$ sample size

$\Sigma X_i \sim \text{Binomial}(n, p)$

$$P(X_1, \cdots X_n \mid \Sigma X_i = t)$$

$$= \frac{P(X_1 = x_1, \cdots X_n = x_n, \Sigma X_i = t)}{\binom{n}{t} p^t (1-p)^{n-t}}$$

$$= \begin{cases} 0, & \text{if } \Sigma x_i \neq t \\ \dfrac{P(X_1 = x_1, \cdots X_n = x_n)}{\binom{n}{t} p^t (1-p)^{n-t}}, & \text{if } \Sigma x_i = t \end{cases}$$

i.i.d

$$= \frac{\prod_{i=1}^{n} P(X_i = x_i)}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{\cancel{p^t (1-p)^{n-t}}}{\binom{n}{t} \cancel{p^t (1-p)^{n-t}}} = \frac{1}{\binom{n}{t}}.$$

Therefore, by the def of sufficiency, $\sum_{i=1}^{n} X_i$ is a sufficient statistic for p.

Sufficient statistic - *Directly use the definition*

is very hard to work with.

Example 4 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$. Is $\bar{X}_n$ sufficient for μ ?

Recap: 1. $\mathbf{X} = (X_1, \ldots, X_k)^T$ *is multivariate normal if and only if any linear combination* $Y = a_1 X_1 + \cdots + a_k X_k$ *is normal.*

2. Conditional distribution of multivariate Normal

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_\mathbf{X} & \Sigma_\mathbf{XY} \\ \Sigma_\mathbf{YX} & \Sigma_\mathbf{Y} \end{pmatrix} \right)$$

$\mathbf{Y}|\mathbf{X}=\mathbf{x}$ *is still multivariate Normal with*

$$E(\mathbf{Y}|\mathbf{X}=\mathbf{x}) = \mu_\mathbf{Y} + \Sigma_\mathbf{YX}\Sigma_\mathbf{X}^{-1}(\mathbf{x} - \mu_X)$$

$$\mathrm{Cov}(\mathbf{Y}|\mathbf{X}=\mathbf{x}) = \Sigma_\mathbf{Y} - \Sigma_\mathbf{YX}\Sigma_\mathbf{X}^{-1}\Sigma_\mathbf{XY}$$

Solution:* $(X_1, \ldots, X_n) | \bar{X}_n$

$X \sim N(\ \ )$
$Y \sim N(\ \ )$
$\not\Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix}$

① We need to ensure that

$\begin{pmatrix} \bar{X}_n \\ X_1 \\ \vdots \\ X_n \end{pmatrix}$ is multivariate normal.

iid

$a_0 \bar{X}_n + a_1 X_1 + \cdots + a_n X_n$

$= a_0 \frac{1}{n} \sum_{i=1}^n X_i + \sum_{i=1}^n a_i X_i = \sum_{i=1}^n \left(\frac{a_0}{n} + a_i \right) X_i \sim N(\ ,\ )$

Thus, $\begin{pmatrix} \bar{X}_n \\ X_1 \\ \vdots \\ X_n \end{pmatrix}$ is multivariate normal.

② $E\bar{X}_n = \mu$

$EX_i = \mu$

$\mathrm{cov}(\bar{X}_n, X_i) = \mathrm{cov}\left(\frac{X_1 + \cdots + X_n}{n}, X_i \right) \underset{\text{iid}}{=} \mathrm{cov}\left(\frac{X_i}{n}, X_i \right) = \frac{\sigma^2}{n}$.

$\mathrm{cov}(X_i, X_j) = 0,\ i \neq j,\ \mathrm{cov}(X_i, X_i) = \sigma^2$.

Therefore,

$$\begin{pmatrix} \bar{X}_n \\ X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{n} & \frac{\sigma^2}{n} & \cdots & \frac{\sigma^2}{n} \\ \frac{\sigma^2}{n} & \sigma^2 & 0 & \cdots & 0 \\ \vdots & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \frac{\sigma^2}{n} & 0 & 0 & \cdots & \sigma^2 \end{pmatrix}\right]$$

μ_X, μ_Y, Σ_X, Σ_{XY}, Σ_Y

Denote $\vec{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$(X_1, \dots X_n) \mid \bar{X}_n \sim N\left(\bar{X}_n \vec{1}_n,\ \sigma^2 I - \frac{\sigma^2}{n}\vec{1}_n \vec{1}_n^T\right)$ which does not depend on μ.

$$\mu_Y + \Sigma_{YX}\Sigma_X^{-1}(X - \mu_X) = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} + \begin{pmatrix} \frac{\sigma^2}{n} \\ \vdots \\ \frac{\sigma^2}{n} \end{pmatrix} \left(\frac{\sigma^2}{n}\right)^{-1} (\bar{X}_n - \mu)$$

$$= \mu \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (\bar{X}_n - \mu) = \bar{X}_n \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \bar{X}_n \vec{1}_n$$

$$\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY} = \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} - \begin{pmatrix} \frac{\sigma^2}{n} \\ \vdots \\ \frac{\sigma^2}{n} \end{pmatrix} \left(\frac{\sigma^2}{n}\right)^{-1} \left(\frac{\sigma^2}{n}, \cdots \frac{\sigma^2}{n}\right)$$

$$= \sigma^2 I - \frac{\sigma^2}{n} \vec{1}_n \vec{1}_n^T$$

Therefore, $\bar{X}_n$ is sufficient for μ.

Sufficient statistic - *Work around the definition*

because $(X_1, \cdots X_n) | T(\underline{X}_n)$ is very difficult to derive.

Fisher-Neyman Factorization Theorem: Denote $\mathbf{x}_n = (x_1, \ldots, x_n)$.

Theorem E. Let $f(\mathbf{x}_n \,|\, \theta)$ be the joint pdf/pmf of the sample $\mathbf{X}_n$. Then $T(\mathbf{X}_n)$ is a sufficient statistic *iff* there exists functions $g(t, \theta)$ and $h(\mathbf{x}_n)$ such that

$$\text{for any } \mathbf{x}_n \text{ and } \theta,\ f(\mathbf{x}_n \,|\, \theta) = g[T(\mathbf{x}_n), \theta] \cdot h(\mathbf{x}_n).$$

Example 7 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $\text{Bernoulli}(p)$. Is $\sum_{i=1}^{n} X_i$ sufficient for p?

$$f(x_1, \cdots x_n | p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\Sigma x_i} (1-p)^{1-\Sigma x_i}$$

$$= \left(\frac{p}{1-p}\right)^{\Sigma x_i} (1-p) \cdot 1$$

ΣX_i is sufficient.

Sufficient statistic - *Work around the definition*

$\vec{x}_n$

$\underline{X}_n$

Fisher-Neyman Factorization Theorem: Denote $\mathbf{x}_n = (x_1, \ldots, x_n)$.

Theorem E. Let $f(\mathbf{x}_n \,|\, \theta)$ be the joint pdf/pmf of the sample $\mathbf{X}_n$. Then $T(\mathbf{X}_n)$ is a sufficient statistic *iff* there exists functions $g(t,\, \theta)$ and $h(\mathbf{x}_n)$ such that

$$\text{for any } \mathbf{x}_n \text{ and } \theta,\ f(\mathbf{x}_n \,|\, \theta) = g[T(\mathbf{x}_n), \theta] \cdot h(\mathbf{x}_n).$$

Proof*. We only prove the case when X is discrete, i.e. $f(x_n|\theta)$ is a pmf.

"$\Leftarrow$" If T is a sufficient statistic,

$$P_\theta(X_1 = x_1, \cdots X_n = x_n \,|\, T(\underline{X}_n) = t) = \frac{P_\theta(X_1 = x_1, \cdots X_n = x_n, T(\underline{X}_n) = t)}{P_\theta(T(\underline{X}_n) = t)}$$

By sufficiency, does involve θ.

$$= \begin{cases} 0, & \sum x_i \neq t \\ \dfrac{P_\theta(X_1 = x_1, \cdots X_n = x_n)}{P_\theta(T(\underline{X}_n) = \sum x_i)}, & \sum x_i = t \end{cases}$$

$$P_\theta(X_1 = x_1, \cdots X_n = x_n) = \underbrace{P_\theta(T(\underline{X}_n) = \sum x_i)}_{g(T(\underline{X}_n), \theta)} \cdot \underbrace{P(X_1 = x_1, \cdots X_n = x_n \,|\, T(\underline{X}_n) = t)}_{h(\vec{x}_n)}$$

Sufficient statistic - *Work around the definition*

Proof cont'd. "$\Rightarrow$" If $P_\theta(X_1=x_1, \ldots X_n=x_n) = g(T(\vec{x}_n),\theta)\, h(\vec{x}_n)$

$$P_\theta(X_1=x_1, \ldots X_n=x_n \mid T(\vec{X}_n)=t) = \frac{P_\theta(X_1=x_1, \ldots X_n=x_n, T(\vec{X}_n)=t)}{P_\theta(T(\vec{X}_n)=t)}$$

$$= \begin{cases} 0, & \sum x_i \neq t \\ \dfrac{P_\theta(X_1=x_1, \ldots X_n=x_n)}{P_\theta(T(\vec{X}_n)=t)} = \dfrac{P_\theta(X_1=x_1, \ldots X_n=x_n)}{\sum_{\{x_i : \sum x_i = t\}} P_\theta(X_1=x_1, \ldots X_n=x_n)} & \sum x_i = t \end{cases}$$

$$P_\theta(X_1=x_1, \ldots X_n=x_n) = g(\underbrace{T(\vec{x}_n)}_{t},\theta)\cdot h(\vec{x}_n)$$

$$\to g(\underbrace{T(\vec{x}_n)}_{t},\theta)\, h(\vec{x}_n)$$

$$\frac{\cancel{g(t,\theta)}\, h(\vec{x}_n)}{\sum_{\{x_i : \sum x_i = t\}} \cancel{g(t,\theta)}\, h(\vec{x}_n)} = \frac{h(\vec{x}_n)}{\sum_{\{x_i : \sum x_i = t\}} h(\vec{x}_n)}$$

Only depends on $\vec{x}_n$.

$\Rightarrow$ sufficiency ∎

Sufficient statistic - *Work around the definition*

Sufficiency : for any $\mathbf{x}_n$ and θ, $f(\mathbf{x}_n \mid \theta) = g[T(\mathbf{x}_n), \theta] \cdot h(\mathbf{x}_n)$.

Corollary E. If $T(\mathbf{X}_n)$ is sufficient for θ, the maximum likelihood estimate is a function of $T(\mathbf{X}_n)$.

$$\arg\max_{\theta} f(\vec{x}_n \mid \theta) = \arg\max_{\theta} g(T(\vec{x}_n), \theta)\, \underbrace{h(\vec{x}_n)}_{}$$

→ does not vary with θ

$$= \underbrace{\arg\max_{\theta} \underbrace{g(T(\vec{x}_n), \theta)}_{}}_{}$$

Therefore, the argmax should be a function of $T(\vec{x}_n)$.

Sufficient statistic - *Work around the definition*

indicator func $\to \mathbb{1}\{\text{condition}\} = \begin{cases} 1, & \text{if cond is true} \\ 0, & \text{if cond is not} \end{cases}$

Example 8. Let $X_1, \ldots, X_n$ be i.i.d $U(-\theta, \theta)$. Is the MLE for θ sufficient?

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

$$f(\vec{x}_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \left(\frac{1}{2\theta}\right)^n \Leftarrow \theta \geq \max\{-X_{(1)}, X_{(n)}\}$$

$\downarrow$

likelihood

$$= 0 \Leftarrow \theta < \max\{-X_{(1)}, X_{(n)}\}$$

$$\Leftrightarrow \quad f(\vec{x}_n|\theta) = \left(\frac{1}{2\theta}\right)^n \mathbb{1}\left(\theta \geq \max\{-X_{(1)}, X_{(n)}\}\right)$$

$$g(\hat{\theta}_{MLE}, \theta) \cdot 1 \Rightarrow \hat{\theta}_{MLE} \text{ is sufficient for } \theta.$$

Sufficient statistic - *Work around the definition*

Example 4 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$. Is $\bar{X}_n$ sufficient for μ? Can you find a sufficient statistic for (μ, σ^2)?

Solution. Recall $f(\mathbf{x}_n \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$.

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2+\mu^2-2x_i\mu}$$

$c_2(\mu, \sigma^2)$

$$= \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum x_i^2}}_{h(\bar{x}_n)} \underbrace{e^{-\frac{n\mu^2}{2\sigma^2}}}_{c(\mu)} \underbrace{e^{\frac{n\mu\frac{\sum x_i}{n}}{\sigma^2}}}_{h(\sum x_i,\ \mu)}$$

$c_1(\sigma^2)$

$h'(\bar{X}_n, \mu)$

If we are finding sufficient statistic for μ only, then we can σ^2 as a known parameter.

If we are finding a sufficient statistic for both μ & σ^2, $(\sum x_i^2, \sum x_i)$ are together sufficient.

$\left.\begin{array}{l}\sum x_i \\ \bar{X}_n\end{array}\right\} \Rightarrow$ sufficient

Sufficient statistic - *Work around the definition*

Example 9. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu,\, \mu^2)$. Is $\bar{X}_n$ sufficient for μ?

Solution. Now $f(\mathbf{x}_n \mid \mu,\, \mu^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\mu^2}} e^{-\frac{1}{2\mu^2}(x_i-\mu)^2}$.

$$= \underbrace{\left(\frac{1}{\sqrt{2\pi\mu^2}}\right)^n}_{c_1(\mu)} \underbrace{e^{-\frac{1}{2\mu^2}\Sigma x_i^2}}_{g_1(\Sigma x_i^2,\, \mu)} \underbrace{e^{-\frac{n}{2}}}_{c_2} \underbrace{e^{\frac{\Sigma x_i}{\mu}}}_{g_2(\Sigma x_i,\, \mu)}$$

To be sufficient for μ, we have to include both Σx_i^2 and Σx_i.

***Two** statistics are needed for **one** parameter sometimes!*

Exponential family

Definition. The **exponential family** of distributions has pdf/pmf of the form

$$f(x \mid \theta) = h(x) c(\theta) \exp\left[\sum_{i=1}^{k} w_i(\theta) t_i(x)\right].$$

Many common distributions, including the normal, the binomial, the Poisson, and the gamma, are members of this family.

In this case, the joint density of $X_1, \ldots, X_n$ is:

$$f(\mathbf{x}_n \mid \theta) = \left[\prod_{j=1}^{n} h(X_i)\right] c(\theta) \exp\left\{\sum_{i=1}^{k} w_i(\theta)\left[\sum_{j=1}^{n} t_i(X_j)\right]\right\}.$$

$$\prod_{i=1}^{n} f(x_i \mid \theta)$$

Therefore, the sufficient statistic for θ

is $\left(\sum_{j=1}^{n} t_1(x_j), \;\ldots,\; \sum_{j=1}^{n} t_k(x_j) \right)$.

$\theta = (\theta_1, \ldots, \theta_d)$

d=k, full exponential family

d<k, curved exponential family

d>k, not seen.

Exponential family

Example 10. Gamma distribution is a member of the exponential family.

$$f(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Full exponential.

$$= \frac{\beta^2}{\Gamma(\alpha)} e^{(\alpha-1)\log x} e^{-\beta x}$$

$$\downarrow$$

$$C(\alpha, \beta)$$

$$= \frac{\beta^2}{\Gamma(\alpha)} e^{\underbrace{(\alpha-1)}\overbrace{\log x}^{t_1(x)} \underbrace{-\beta}\overbrace{x}^{t_2(x)}}$$

$\Rightarrow$ The sufficient statistic for (α, β) should be

$$\left(\sum_{i=1}^n \log X_i , \sum_{i=1}^n X_i \right).$$

$\boxtimes$

$E(E(X|Y)) = EX$

$\hat{\delta}_n \to \delta(\theta)$

$\downarrow T$

Sufficiency helps us find UMVUE

Rao-blackwell Theorem: $\mathrm{var}\, X = \mathrm{Var}(E(X|Y)) + E[\mathrm{var}(X|Y)]$

$\phi(T)$ better, more efficient statistic

Theorem F. Let $\hat{\delta}_n$ be any unbiased estimator of $\delta(\theta)$ and let $T(\mathbf{X}_n)$ be a sufficient statistic for θ. Define a new statistic $\phi(T) = E\left(\hat{\delta}_n \middle| T\right)$. Then we have

$$E_\theta[\phi(T)] = \delta(\theta) \text{ and } \mathrm{Var}_\theta[\phi(T)] \le \mathrm{Var}_\theta\left[\hat{\delta}_n\right], \forall\theta.$$

Remark: To find UMVUE, we can simply restrict our attention to estimators that are functions of $T(\mathbf{X}_n)$.

Proof. $E_\theta[\phi(T)] = E_\theta(E_\theta(\hat{\delta}_n|T)) = E_\theta \hat{\delta}_n = \delta(\theta)$.

$\mathrm{var}_\theta(\hat{\delta}_n) = \mathrm{var}_\theta(\underbrace{E(\hat{\delta}_n|T)}_{=\phi(T)}) + E_\theta(\underbrace{\mathrm{var}(\hat{\delta}_n|T)}_{\geq 0})$

$= \mathrm{var}_\theta\, \phi(T) + E_\theta(\mathrm{var}(\hat{\delta}_n|T)) \geq 0$

$\geq \mathrm{var}_\theta\, \phi(T)$.

Evaluating estimators: a summary

~~δ_n~~ $T \to$ sufficient & complete

$\to \phi(T)$ must be the unique UMVUE.

- $\bar{X}_n$ and $\hat{\sigma}_n^2$
- MM $\in$
- MLE $\in$

When n is small,

$$\text{MSE} = \text{variance} + \underbrace{\text{bias}^2}_{}$$

$0 =$ bias

~~Minimal MSE~~

UMVUE $\leftarrow$

~~CR bound~~

Rao-blackwell
+ sufficiency

Lehmann-Scheffé
+completeness

When n is sufficiently large,

Consistency

Asymptotic normality

Exact distribution of $\bar{X}_n$ and $\hat{\sigma}_n^2$ under $N(\mu, \sigma^2)$

8.5.3 of Rice

07/01/2021

Exact sampling distribution under $N(\mu, \sigma^2)$

Theorem G. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \mu^2)$. Then $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, $\sigma^2 \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2 \sim \chi^2_{n-1}$ and they are *independent* of each other.

Proof*.

Exact sampling distribution under $N(\mu, \sigma^2)$

Definition. A student *t* distributed r.v. with df=n can be generated using independent $Z \sim N(0,1)$ and $U \sim \chi^2_n$:

$$\frac{Z}{\sqrt{U/n}} \sim t_n.$$

$$\Rightarrow \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Next Tuesday ...

- Exact CI for μ and σ^2 under $N(\mu, \sigma^2)$;
- Hypothesis testing.