Mathematical Statistics2 Tutoring5

Seung Bong Jung

Seoul National University

November 1, 2021

Rao-Blackwell Theorem

- Rao-Blackwell theorem: Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot;\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. Let $Y = u(X)$ be a sufficient statistic. Then, for any estimator $\hat{\eta}(X)$ of $\eta = g(\theta)$ with finite second moment, $\hat{\eta}^* = E(\hat{\eta}(X)|Y)$ is a statistic with the properties that (1) $E_\theta(\hat{\eta}^*) = E_\theta(\hat{\eta})$, (2) $\text{var}_\theta(\hat{\eta}^*) \leq \text{var}_\theta(\hat{\eta})$, (3) $\text{MSE}_\theta(\hat{\eta}^*) \leq \text{MSE}_\theta(\hat{\eta})$.
- The Rao-Blackwell theorem tells that one may have a better estimator by conditioning on a sufficient statistic, n terms of MSE. By the theorem, if $\hat{\eta}$ is an unbiased estimator of η, then $\hat{\eta}^*$ is also an unbiased estimator but with a smaller variance.
- Proof of the theorem: $\text{var}(W) = E(\text{var}(W|V)) + \text{var}(E(W|V))$.

Example: Rao-Blackwellization

Let $X_1, \dots, X_n$ ($n \geq 2$) be a random sample from $U[0, \theta]$, $\theta > 0$. Take $\hat{\theta} = 2\bar{X}$ as an unbiased estimator of θ. We know that $X_{(n)}$ is a sufficient statistic for $\theta > 0$. By Rao-Blackwell theorem, $\hat{\theta}^* \equiv E(2\bar{X}|X_{(n)})$ is an UE of θ with variance less than or equal to that of $\hat{\theta}$. For $1 \leq r \leq n-1$, we note that

$$\text{pdf}_{X_{(r)}|X_{(n)}}(x|y) = \frac{(n-1)!}{(r-1)!(n-r-1)!}(\frac{x}{y})^{r-1}(1-\frac{x}{y})^{n-r-1}\frac{1}{y}I_{(0,y)}(x)$$

and that, recalling the pdf of Beta$(r+1, n-r-1)$,

$$\int_0^y x \cdot \frac{(n-1)!}{(r-1)!(n-r-1)!}(\frac{x}{y})^{r-1}(1-\frac{x}{y})^{n-r-1}\frac{1}{y}dx = (r/n)y.$$

Example: Rao-Blackwellization

Thus, we get

$$
\begin{aligned}
2E(\bar{X}|X_{(n)} = y) &= 2n^{-1}\left(y + \sum_{r=1}^{n-1} E(X_{(r)}|X_{(n)} = y)\right) \\
&= 2n^{-1}(y + \sum_{r=1}^{n-1} \frac{r}{n} \cdot y) \\
&= \frac{n+1}{n} y.
\end{aligned}
$$

Indeed, $\hat{\theta}^* = (n+1)X_{(n)}/n$ and

$$
\begin{aligned}
\mathrm{var}_\theta(\hat{\theta}^*) &= (\frac{n+1}{n})^2 \cdot \mathrm{var}_\theta(X_{(n)}) \\
&= \frac{1}{n(n+2)}\theta^2 < \frac{1}{3n}\theta^2 = \mathrm{var}_\theta(\hat{\theta}).
\end{aligned}
$$

Uniformly Minimum Variance Unbiased Estimator

We have seen that taking conditional expectation, on a sufficient statistic, of a given unbiased estimator always improves the estimator in terms of variance.

- UMVUE: An estimator $\hat{\eta}$ of $\eta = g(\theta)$ is called the uniformly minimum variance unbiased estimator if it itself is unbiased and $\text{var}_\theta(\hat{\eta}) \leq \text{var}_\theta(\tilde{\eta})$ for all $\theta \in \Theta$ and for any unbiased estimator $\tilde{\eta}$ of η.

Complete Statistic

Let $X_1, \ldots, X_n$ be a random sample from a population with p.d.f. $f(\cdot;\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. The following notion of completeness facilitates the derivation of UMVUE.

- Complete statistic for $\theta \in \Theta$: A statistic $Y = u(X)$ is called a complete statistic for $\theta \in \Theta$ and $\{\mathrm{pdf}_Y(\cdot;\theta) : \theta \in \Theta\}$ is called a complete family of distributions if

$$E_\theta \phi(Y) = 0 \text{ for all } \theta \in \Theta \text{ implies } P_\theta(\phi(Y) = 0) = 1 \text{ for all } \theta \in \Theta.$$

- A complete statistic Y is complete in sense that any non-constant function of Y has a non-constant expected value (as a function of θ).
- Complete sufficient statistic: A statistic is called a complete sufficient statistic (CSS) for $\theta \in \Theta$ if it is sufficient and complete for $\theta \in \Theta$.

Rao-Blackwell-Lehmann-Scheffe Theorem

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot;\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. Let $Y = u(X)$ be a CSS for $\theta \in \Theta$. Assume that there exists an unbiased estimator with finite variance.

- For any unbiased estimator $\hat{\eta}_0$ of $\eta = g(\theta)$ with finite variance, $\hat{\eta} = E(\hat{\eta}_0|Y)$ is the UMVUE
- Any function of Y, say $\phi(Y)$, is the UMVUE if it is unbiased.

Methods of Finding UMVUE

- Method 1: Rao-Blackwellization with a CSS
  (1) Find a CSS $Y = u(X)$.
  (2) Find an easy UE $\hat{\eta}_0$.
  (3) compute the conditional expectation $E(\hat{\eta}_0|Y)$.
- Method 2: Trial and error
  (1) Find a CSS $Y = u(X)$. 2) Solve $E_\theta \phi(Y) \overset{\theta}{=} g(\theta)$ with respect to ϕ, or try some $\phi(Y)$ and check unbiasedness.

CSS and UMVUE: Uniform $[-\theta, \theta]$ Model

- Let $X_1, \ldots, X_n$ $(n \geq 2)$ be a random sample from Uniform$[-\theta, \theta]$, $\theta > 0$.
- In this model, $Y = \max_{1 \leq i \leq n} |X_i|$ is a CSS and $\hat{\eta} = (n+1)/n \cdot Y$ is the UMVUE of θ.

Ancillary Statistic

- Let $X_1, \dots, X_n$ $(n \geq 2)$ be a random sample from a population with pdf $f(\cdot;\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$.
- Ancillary statistic: A statistic $Z = v(X)$ is called an ancillary statistic for $\theta \in \Theta$ if $P_\theta(Z \in A)$ does not depend on $\theta \in \Theta$ for all A.
- Basu's Theorem (Independence of CSS and AS): If $Y = u(X)$ is a CSS and $Z = v(X)$ is an AS for $\theta \in \Theta$, then Y and Z are independent under P_θ for all $\theta \in \Theta$.

Ancillary Statistic: Examples

- N(θ,1), $\theta \in \mathbb{R}$:

$$(X_1 - \bar{X}, \ldots, X_n - \bar{X}) \stackrel{d}{\equiv} (Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z})$$

for Z_i being i.i.d. from $N(0, 1)$.

- Exp(θ,1), $\theta \in \mathbb{R}$:

$$(X_1 - X_{(1)}, \ldots, X_n - X_{(n)}) \stackrel{d}{\equiv} (Z_1 - Z_{(1)}, \ldots, Z_n - Z_{(n)})$$

for Z_i being i.i.d. from Exp$(0, 1)$.

- Gamma(α, β), $\beta > 0$ with α known:

$$\left(\frac{X_1}{\sum_{i=1}^{n+1} X_i}, \cdots, \frac{X_n}{\sum_{i=1}^{n+1} X_i} \right) \stackrel{d}{\equiv} \left(\frac{Z_1}{\sum_{i=1}^{n+1} Z_i}, \cdots, \frac{Z_n}{\sum_{i=1}^{n+1} Z_i} \right)$$

for Z_i being i.i.d. from Gamma$(\alpha, 1)$.

Ancillary Statistic: Examples

- Uniform$(0,\theta)$, $\theta > 0$: $X_{(n)}/X_{(1)} \overset{d}{\equiv} Z_{(n)}/Z_{(1)}$ for Z_i being i.i.d. from Uniform$(0,1)$.
- N(μ,σ^2), $(\mu\sigma^2) \in \mathbb{R} \times \mathbb{R}_+$:

$$\left(\frac{X_1-\bar{X}}{S_X},\dots,\frac{X_n-\bar{X}}{S_X}\right) \overset{d}{\equiv} \left(\frac{Z_1-\bar{Z}}{S_Z},\dots,\frac{Z_n-\bar{Z}}{S_Z}\right)$$

  for Z_i being i.i.d. from $N(0,1)$, where $S_X^2 = \sum_{i=1}^n (X_i-\bar{X})^2/(n-1)$ and $S_Z^2 = \sum_{i=1}^n (Z_i-\bar{Z})^2/(n-1)$.
- Exp(μ,σ), $(\mu,\sigma) \in \mathbb{R} \times \mathbb{R}_+$:

$$\frac{X_1-X_{(1)}}{\sum_{i=1}^n (X_i-X_{(1)})} \overset{d}{\equiv} \frac{Z_1-Z_{(1)}}{\sum_{i=1}^n (Z_i-Z_{(1)})}$$

  for Z_i being i.i.d. from Exp$(0,1)$.

Basu's Theorem: Examples

- Let $X_1, \ldots, X_n$ be a random sample from Exp(μ, σ), $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$. Then, $X_{(1)}$ and $\sum_{i=1}^{n}(X_i - X_{(1)})$ are independent.
- Ket $X_1, \ldots, X_{n+1}$ be a random sample from Gamma(α, β), $\alpha > 0, \beta > 0$. Then, $\sum_{i=1}^{n+1} X_i$ is independent of

$$Z \equiv \left(\frac{X_1}{\sum_{i=1}^{n+1} X_i}, \cdots, \frac{X_n}{\sum_{i=1}^{n+1} X_i} \right).$$

Exponential Family

A family of distributions $\{f(\cdot;\theta) : \theta \in \Theta\}$ for $\Theta \subset \mathbb{R}^d$ is called exponential family if
(1) the support of the density $f(\cdot;\theta)$ does not depend on $\theta \in \Theta$. (2) the density has the following form:

$$f(x;\theta) = \exp(\eta(\theta)^t T(x) - B(\theta)) \cdot h(x)$$

for some known functions $\eta = (\eta_1, \ldots, \eta_k)^t$, $T = (T_1, \ldots, T_k)^t$, B and h.
An exponential family is called $k-$parameter regular exponential family if
(3) $\eta(\Theta) \equiv \{\eta(\theta) : \theta \in \Theta\} \subset \mathbb{R}^k$ contains a $k-$dimensional open rectangle.

Random Sample from Exponential Family

Let $X_1, \dots, X_n$ be a random sample from an exponential family of pdf's $f(\cdot;\theta)$, $\theta \in \Theta$ with $f(x;\theta) = \exp(\eta(\theta)^t T(x) - B(\theta)) \cdot h(x)$, where $\eta(\theta)$ is a $k-$vector. Let $\mathcal{X}$ denote the common support of $f(\cdot;\theta)$. Then,
(1) the joint densities of $(X_1, \dots, X_n)$ also form an exponential family with $\mathcal{X}^n$ as the common support and

$$\prod_{i=1}^{n} f(x_i;\theta) = \exp(\eta(\theta)^t \sum_{i=1}^{n} T(x_i) - nB(\theta)) \cdot \prod_{i=1}^{n} h(x_i)$$

as the joint density;
(2) If $\eta(\Theta)$ contains a $k-$dimensional open rectangle, then $Y = \sum_{i=1}^{n} T(X_i)$ is a CSS for $\theta \in \Theta$.

MGF/CGF of $T(X)$ in Exponential Family

Let X be a random variable having a pdf $f(\cdot, \eta)$, $\eta \in \mathcal{N} \subset \mathbb{R}^k$. Assume $f(x;\eta) = \exp(\eta^t T(x) - A(\eta)) \cdot h(x)$ and that $\mathcal{N}$ contains a $k-$dimensional open rectangle. Then,

(3) the cumulant generating function of $T(X)$ is given by

$$\text{cgf}_{T(X)}(u;\eta) \equiv \log E_\eta e^{u^t T(x)} = A(\eta + u) - A(\eta)$$

for all $\eta \in \text{Int}(\mathcal{N})$.

(4) the mean and variance of $T(X)$ under P_η with $\eta \in \text{Int}(\mathcal{N})$ are then given by

$$E_\eta T(X) = \dot{A}(\eta), \text{var}_\eta(T(X)) = \ddot{A}(\eta).$$

Let X be a random variable having a pdf $f(\cdot, \eta)$, $\eta \in \mathcal{N} \subset \mathbb{R}^k$. Assume $f(x;\eta) = \exp(\eta^t T(x) - A(\eta)) \cdot h(x)$ and that $\mathcal{N}$ contains a $k-$dimensional open rectangle. Then,
(5) The log-likelihood is strictly concave, and the unique MLE of η is determined by the likelihood equation

$$n^{-1} \sum_{i=1}^{n} T(X_i) = \dot{A}(\eta),$$

provided that it has a solution $\hat{\eta} \in \mathcal{N}$.

Multinomial Experiments

Let $X_i = (X_{i,1}, \ldots, X_{i,k-1})^t$ be i.i.d. Multinomial$(1, p)$, $p \equiv (p_1, \ldots, p_{k-1})^t$, $p_j > 0, p_1 + \cdots + p_{k-1} < 1$.

- Let $p_k = 1 - p_1 - \cdots - p_{k-1}$. Then, the common density of X_i is given by

$$f(x; p) = \exp(x_1 \log(p_1/p_k) + \cdots + x_{k-1} \log(p_{k-1}/p_k) + \log p_k),$$

  so that the distributions of X_i form a $(k-1)-$parameter regular exponential family.

- $Y = \sum_{i=1}^n X_i = (\sum_{i=1}^n X_{i,1}, \ldots, \sum_{i=1}^n X_{i,k-1})^t$ is a CSS for p.
- MLE of η: The MLE of $\eta \equiv (\log(p_1/p_k), \ldots, \log(p_{k-1}/p_k))^t \stackrel{\text{let}}{=} h(p)$ solves the equation

$$Y/n = E_\eta X_1, \text{i.e., } Y/n = h^{-1})(\eta).$$

  Thus, the MLE of η is given by $\hat{\eta} = h(Y/n)$.

Multinomial Experiments

- MLE of p: The MLE of p is then $\hat{p} = h^{-1}(\hat{\eta}) = Y/n$.
- UMVUE of p: Y/n is an UE of p and is a function of the CSS Y, so that $\hat{p} = Y/n$ is also the UMVUE of p.
- UMVUE of $\Sigma \equiv \mathrm{diag}(p) - pp^t$: Here, an estimator $\hat{\Sigma}$ of Σ is called the UMVUE of Σ if $\mathrm{var}_p(\hat{\Sigma}^{\mathrm{UE}}) - \mathrm{var}_p(\hat{\Sigma})$ is nonnegative definite for all $\hat{\Sigma}^{\mathrm{UE}}$ and for all p, with $\hat{\Sigma}$ and $\hat{\Sigma}^{\mathrm{UE}}$ being the vectorized versions. Note that

$$\hat{\Sigma}^{\mathrm{MLE}} = \mathrm{diag}(Y/n) - (Y/n)(Y/n)^t.$$

Computing the expected value of $\hat{\Sigma}^{\mathrm{MLE}}$, we get

$$\begin{aligned} E_p(\hat{\Sigma}^{\mathrm{MLE}}) &= \mathrm{diag}(p) - \mathrm{var}_p(Y/n) - E_p(Y/n)E_p(Y/n)^t \\ &= \mathrm{diag}(p) - n^{-1}\Sigma - pp^t = (1 - 1/n)\Sigma. \end{aligned}$$

Thus, $\hat{\Sigma}^{\mathrm{UMVUE}} = n/(n-1) \cdot \hat{\Sigma}^{\mathrm{MLE}}$.

Multivariate Normal Population

Let $X_i = (X_{i,1}, \ldots, X_{i,k})^t$ $(n \geq 2)$ be i.i.d. Normal$(\mu, \Sigma), \mu \in \mathbb{R}^k$ and Σ in the set of $k \times k$ positive definite matrices.

- With $\theta \equiv (\mu, \Sigma) \in \mathbb{R}^d$ for $d = k + k(k+1)/2$,

$$\begin{aligned} f(x;\theta) &= |2\pi\Sigma|^{-1/2} \exp(-(x-\mu)^t \Sigma^{-1}(x-\mu)/2) \\ &= \exp(-\text{tr}(\Sigma^{-1}xx^t)/2 + \mu^t\Sigma^{-1}x - \mu^t\Sigma^{-1}\mu/2 - 1/2 \log|2\pi\Sigma|). \end{aligned}$$

- $Y = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i X_i^t)$ is a CSS for θ, where $\sum_{i=1}^n X_i X_i^t$ is understood to be a $k(k+1)/2-$vector.
- MLE of $\eta \equiv (\Sigma^{-1}\mu, \Sigma^{-1})$: It is the solution of

$$Y/n = E_\eta(X_1, X_1X_1^t), \text{i.e., } Y/n = (\mu, \Sigma + \mu\mu^t) \stackrel{\text{let}}{=} g(\mu, \Sigma).$$

Let $h(\mu, \Sigma) = (\Sigma^{-1}\mu, \Sigma^{-1})$. Then, $\hat{\eta}^{\text{MLE}} = h \circ g^{-1}(Y/n)$.

Multivariate Normal Population

- MLE of μ and Σ: The MLE of $(\mu, \Sigma) = h^{-1}(\eta)$ is then

$$(\hat{\mu}^{\mathsf{MLE}}, \hat{\Sigma}^{\mathsf{MLE}}) = h^{-1} \circ h \circ g^{-1}(Y/n) = g^{-1}(Y/n).$$

  By the definition of the function $g : \mathbb{R}^d \to \mathbb{R}^d$, this means

$$n^{-1} \sum_{i=1}^{n} X_i = \hat{\mu}^{\mathsf{MLE}} \text{ and } n^{-1} \sum_{i=1}^{n} X_i X_i^t = \hat{\Sigma}^{\mathsf{MLE}} + \hat{\mu}^{\mathsf{MLE}} \hat{\mu}^{\mathsf{MLE}t}$$

  so that $\hat{\mu}^{\mathsf{MLE}} = \bar{X}$ and $\hat{\Sigma}^{\mathsf{MLE}} = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^t$.
- UMVUE of μ and Σ: Since $(\bar{X}, (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^t)$ is a $1-1$ function of Y, it is also a CSS for (μ, Σ). Since $\bar{X}$ and $(n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^t$ are UEs of μ and Σ, respectively, they are the UMVUE of the respective parameters.