More on UMVUE; Bayes Estimators

Tiandong Wang

Department of Statistics
Texas A&M University

Constructing UMVUE using Rao-Blackwell Method

Now we learn an important method of finding/constructing UMVUEs with the help of complete and sufficient statistics.
Review on conditional expectation:

- (1) $E(X) = E[E(X|Y)]$, for any X, Y.
- (2) $Var(X) = Var[E(X|Y)] + E[Var(X|Y)]$, for any X, Y
- (3) $E(g(X)|Y) = \int g(x) f_{x|y}(x|y) dx$, and it is a function of Y
- (4) If X and Y are independent, then $E(g(X)|Y) = E(g(X))$.

Constructing UMVUE using Rao-Blackwell Method: Example

Examples:

- $Cov(E(X|Y), Y) = Cov(X, Y)$.
- If $X_1, \cdots, X_n$ iid $Pois(\lambda)$, let $Y = \sum_{i=1}^{n} X_i$, then $X_1|Y \sim Bin(Y, \frac{1}{n})$, and $(X_1, \ldots, X_{n-1})|Y \sim \text{Multinomial}(Y; \frac{1}{n}, \ldots, \frac{1}{n}, \frac{1}{n})$

Rao-Blackwell Theorem

Theorem

Let W be unbiased for $\tau(\theta)$ and T be a sufficient statistic for θ. Define $\phi(T) = E(W|T)$. Then:

- *(i) $E_\theta \phi(T) = \tau(\theta)$*
- *(ii) $Var_\theta \phi(T) \leq Var_\theta W$ for all θ.*

Thus, $E(W|T)$ is a uniformly better unbiased estimator of $\tau(\theta)$ than W.

- Remark:
  Conditioning any unbiased estimator on a sufficient statistic will result in a uniform improvement, so we only consider statistics that are functions of a sufficient statistic for best unbiased estimators.
- Examples:
  - For $X_1, \cdots, X_n$ iid Bernoulli(p), show $X_1 X_2$ is unbiased for p^2 but $E(X_1 X_2 | \sum_i X_i)$ is uniformly better.
  - Let $X_1, \cdots, X_n$ be iid $Unif(0, \theta)$. Show $Y = (n+1)X_{(1)}$ is unbiased for θ and $E(Y|X_{(n)})$ is uniformly better.

Uniqueness of UMVUE

Theorem

If W is a best unbiased estimator of $\tau(\theta)$, then W is unique.

Theorem

If $E_\theta W = \tau(\theta)$, W is the best unbiased estimator of $\tau(\theta)$ if and only if W is uncorrelated with all unbiased estimators of 0.

Uniqueness of UMVUE: Examples

Examples:
Let X be an observation from a $Unif(\theta, \theta+1)$. Show that

- (i) $X - 1/2$ is unbiased for θ
- (ii) Show unbiased estimators of zero are periodic functions with period 1. One example is $h(X) = sin(2\pi X)$.
- (iii) Show $X - 1/2$ and $h(X)$ are correlated. So $X - 1/2$ is not the best

Uniqueness of UMVUE: Complete Statistics

Theorem

Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.

Uniqueness of UMVUE: Examples

- Procedure to find the best unbiased estimator of $\tau(\theta)$:
  - (i) Find a complete sufficient statistic T for a parameter θ
  - (ii) Guess a function $\phi(T)$ such that $E[\phi(T)] = \tau(\theta)$
    or construct $\phi(T)$: find an unbiased estimator $h(X)$ of $\tau(\theta)$, then $\phi(T) = E(h(X)|T)$ is the best unbiased estimator of $\tau(\theta)$.
- Example: Suppose $X_1, \cdots, X_n$ are iid Bernoulli(p). Find the best unbiased estimator for p and p^2.

UMVUE: Examples

- Example:Suppose that the random variables $Y_1, \cdots, Y_n$ satisfy $Y_i = \beta x_i + \epsilon_i, i = 1, \cdots, n$, where $x_1, \cdots, x_n$ are fixed constants, and $\epsilon_1, \cdots, \epsilon_n$ are iid $N(0, \sigma^2)$ with σ^2 known. Find the MLE of β and show it is UMVUE.

Summary

Suppose $T = T(\mathbf{X})$ is complete for θ. Then

1. For any parameter $\tau(\theta)$, there is at most one unbiased estimator which is a function of T.
2. If $S = S(\mathbf{X})$ is unbiased for $\tau(\theta)$ and $Var_\theta(S) < \infty$ for all θ, then
$$S^*(\mathbf{X}) = S^* = \mathbb{E}(S|T)$$
is the UMVUE for $\tau(\theta)$.
3. An unbiased estimator with finite variance which is a function of T is the UMVUE.

UMVUE: Examples

Example: Suppose $X_1, \cdots, X_n$ are iid from $N(\mu, \sigma^2)$, both (μ, σ^2) unknown.

- (i) Find the UMVUE for μ.
- (ii) Find the UMVUE for σ^2.
- (iii) Find the UMVUE for μ^2. Is Cramér-Rao bound attained here?

Remarks

Estimating a parameter by UMVUE is one approach to estimation, but may not be very good.

- No unbiased estimator exists.
  - $X_1, \cdots, X_n$ iid Poisson(λ). No unbiased estimator for $1/\lambda$.
- The only one that exists is bad.
  - $X_1, \cdots, X_n$ are iid from $N(\mu, \sigma^2)$, both (μ, σ^2) unknown. The UMVUE for μ^2 can be negative!

Loss Function Optimality

Observations $X_1, \cdots, X_n$ are iid with pdf $f(x|\theta)$, $\theta \in \Theta$. To evaluate the estimator $\hat{\theta}(X)$, various loss function can be used. The loss function measures the closeness of θ and $\hat{\theta}$.

- squared error loss: $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$
- absolute error loss:$L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$
- a loss that penalizes overestimation more than underestimation is

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 I(\hat{\theta} < \theta) + 10(\theta - \hat{\theta})^2 I(\hat{\theta} \geq \theta)$$

- a loss that penalized more if θ is near 0 than if $|\theta|$ is large $L(\theta, \hat{\theta}) = \dfrac{(\theta - \hat{\theta})^2}{|\theta| + 1}$

Loss Function Optimality

- To compare estimators, we use the expected loss, called the risk function,

$$R(\theta, \hat{\theta}) = E_\theta L(\theta, \hat{\theta}(X)).$$

- If $R(\theta, \hat{\theta}_1) < R(\theta, \hat{\theta}_2)$ for all $\theta \in \Theta$, then $\hat{\theta}_1$ is the preferred estimator because it performs better for all θ.
- In particular, for the squared error loss, the risk function is the MSE.

- Example: Suppose $X_1, \cdots, X_n$ are iid from $Bin(1, p)$.Compare two estimators in terms of their MSE.
  - (1) MLE$= \bar{X}$
  - (2) Bayes estimator: prior $\pi(p) \sim Beta(\alpha, \beta)$ with $\alpha = \beta = \sqrt{n/4}$,

$$\hat{p}_B = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}$$

- Example: Suppose $X_1, \cdots, X_n$ are iid from $N(\mu, \sigma^2)$. Consider the estimators of the form $\delta_b(\mathbf{X}) = bS^2$.

An estimator $\widehat{\theta}(\mathbf{X})$ is **inadmissible** wrt the loss function $L(\theta, \hat{\theta})$ if there exists $\widetilde{\theta}(\mathbf{X})$ such that

- $R(\theta, \widetilde{\theta}) \leq R(\theta, \widehat{\theta})$ for all θ and
- $R(\theta, \widetilde{\theta}) < R(\theta, \widehat{\theta})$ for some θ.

In other words, $\widehat{\theta}$ is dominated by some other rule.

Example: X_i iid $N(\mu, \sigma^2)$:

1. If $\mu = 0$ known but σ^2 unknown, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \qquad \text{is inadmissible.}$$

2. If μ, σ^2 unknown, then

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad \text{is inadmissible.}$$

Remarks:

- Admissibility is a critical property, if we believe our loss function.
- Not all admissible estimators are reasonable: in many models a constant estimator $\widehat{\theta}(\mathbf{X}) = 1$ (say) is admissible simply because it has zero risk at $\theta = 1$.
- Bayes estimators are admissible under some regularity conditions.

Let us first see what the Bayes risk is...

Bayes Rule

- The Bayes risk is the average risk with respect to the prior π
$$\int_{\Theta} R(\theta, \hat{\theta})\pi(\theta)d\theta$$
- By definition, the Bayes risk can be written as
$$\int_{\Theta} R(\theta, \hat{\theta})\pi(\theta)d\theta = \int_{\Theta}(\int_{\mathcal{X}} L(\theta, \hat{\theta}(x))f(x|\theta)dx)\pi(\theta)d\theta$$
- Note $f(x|\theta)\pi(\theta) = \pi(\theta|\mathbf{x})m(\mathbf{x})$, where $\pi(\mathbf{x}|\theta)$ is the posterior distribution of θ and $m(\mathbf{x})$ is the marginal distribution of $\mathbf{X}$, then the Bayes risk become
$$\int_{\Theta} R(\theta, \hat{\theta})\pi(\theta)d\theta = \int_{\mathcal{X}}(\int_{\Theta} L(\theta, \hat{\theta}(x))\pi(\theta|x)d\theta)m(\mathbf{x})dx$$
- To minimize the Bayes risk, we only need to find $\hat{\theta}$ to minimize the posterior expected loss for each $\mathbf{x}$.

Bayes Estimator

- The Bayes rule with respect to a prior π is an estimator that yields the smallest value of the Bayes risk.
- Two Bayes rules:
  - (1) For squared error loss, the posterior expected loss is

    $$\int_{\Theta} (\theta - \hat{\theta})^2 \pi(\theta|\mathbf{x}) d\theta = E((\theta - \hat{\theta})^2 |\mathbf{x})$$

    therefore the Bayes rule is $E(\theta|x)$.
  - (2) For absolute error loss, the posterior expected loss is $E(|\theta - a||\mathbf{x})$. The Bayes rule is the median of $\pi(\theta|\mathbf{x})$.

Theorem

For a parameter vector $\boldsymbol{\theta}$, suppose $\hat{\boldsymbol{\theta}}_\pi$ is a Bayes estimator having finite Bayes risk:

1. *w.r.t. a prior pdf π that is positive for all $\boldsymbol{\theta}$, and*
2. *the risk function of every estimator $\hat{\boldsymbol{\theta}}$ is a continuous function of $\boldsymbol{\theta}$.*

Then $\hat{\boldsymbol{\theta}}_\pi$ is admissible.