STA732
Statistical Inference

Lecture 05: Rao-Blackwell Theorem

Yuansi Chen

Spring 2023

Duke University

https://www2.stat.duke.edu/courses/Spring23/sta732.01/

Recap from Lecture 04

- V is **ancillary** if its distribution does not depend on θ
- **Completeness + sufficiency** as the ideal notion of optimal data compression. To prove completeness, one usually goes by definition or by identifying exponential family.
- **Basu's theorem** is useful to prove independence between a complete sufficient statistics and an ancillary statistics.

Goal of Lecture 05

1. Convex loss
2. Rao-Blackwell Theorem
3. Uniformly minimum variance unbiased estimator (UMVU)

Chap. 3.6, 4.1-4.2 in Keener or Chap. 1.7, 2.1 in Lehmann and Casella

We are entering the first approach of arguing for “the best” estimator in point estimation: by restricting to a smaller class of estimators!

Convex loss

Definition. Convex set

A set $\mathcal{C} \subseteq \mathbb{R}^p$ is convex if given any two points $x, y \in \mathcal{C}$, for any $\lambda \in [0, 1]$, we have

$$\lambda x + (1 - \lambda) y \in \mathcal{C}$$

Definition. Convex function

A real-valued function f defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^p$ is a convex function if for any two points $x, y \in \mathcal{C}$ and any $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

It is called strictly convex if the above inequality holds strictly for $x \neq y$ and $\lambda \in (0, 1)$.

Jensen's inequality in finite form

For a convex function $f, x_1, \ldots, x_n$ in its domain, and positive weights α_i with $\sum_{i=1}^n \alpha_i = 1$. Then

$$f(\sum_{i=1}^{n} a_i x_i) \leq \sum_{i=1}^{n} a_i f(x_i)$$

proof by induction, omitted

Jensen's inequality in a probabilistic setting

Jensen's inequality in a probabilistic setting

X is an integrable real-valued random variable, f is convex. Then

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

If f is strictly convex, the inequality holds strictly unless X is almost surely constant.

proof see Thm 3.25, remark 3.26 in Keener or Wikipedia

Examples of convex functions

- $x \mapsto 1/x$ is strictly convex on $(0, \infty)$. Then for $X > 0$, we have

$$\frac{1}{\mathbb{E}[X]} \leq \mathbb{E}[1/X]$$

- $x \mapsto -\log(x)$ is strictly convex on $(0, \infty)$. Then for $X > 0$, we have

$$\log(\mathbb{E}[X]) \geq \mathbb{E}\log(X).$$

Convex loss penalizes extra noise to an estimator

Proposition

Suppose the loss $L(\theta, d)$ is convex in d. Let $\delta(X)$ be an estimate of θ. Define $\tilde{\delta}(X) = \delta(X) + \epsilon$, where ϵ is a zero-mean random variable independent of X. Then

$$R(\theta, \tilde{\delta}) \geq R(\theta, \delta)$$

where the risk $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$

Proof idea: tower property + Jensen's inequality

Rao-Blackwell Theorem

Rao-Blackwell Theorem

Thm 3.28 in Keener

Let T be a sufficient statistics for $\mathscr{P} = \{P_\theta : \theta \in \Omega\}$, let δ be an estimator of $g(\theta)$. Define $\eta(T) = \mathbb{E}_\theta[\delta(X) \mid T]$. If $L(\theta, \cdot)$ is convex, then

$$R(\theta, \eta) \leq R(\theta, \delta).$$

where the risk $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$.

Furthermore, if $L(\theta, \cdot)$ is strictly convex, the inequality is strict unless $\delta(X) \overset{\text{a.s.}}{=} \eta(T)$.

Interpretation

For convex loss functions,

1. If an estimator is not just based on sufficient statistics T, we can improve it.
2. The step of constructing $\eta(T) = \mathbb{E}_\theta[\delta(X) \mid T]$ from δ is called Rao-Blackwellization.
3. When discussing optimal estimators, the only estimators of $g(\theta)$ that are worth considering are functions of sufficient statistics T.

Proof of Rao-Blackwell Theorem

See Keener Thm 3.28, apply Jensen

UMVU

Bias

- The bias of an estimate $\delta(X)$ is $\mathbb{E}_\theta[\delta(X) - g(\theta)]$
- We say an estimator δ is unbiased for $g(\theta)$ if

$$\mathbb{E}_\theta[\delta(X)] = g(\theta), \forall \theta \in \Omega.$$

Ex: what is an unbiased estimator of θ for X drawn from a uniform distribution on $(0, \theta)$?

Bias-variance decomposition under squared error loss

Squared error loss:

$$L(\theta, d) = (d - g(\theta))^2$$

Risk decomposition under squared error loss

Risk becomes the mean squared error $R(\theta, \delta) = \mathbb{E}_\theta \left(\delta(X) - g(\theta)\right)^2$

$$\begin{aligned}
&\mathbb{E}_\theta \left(\delta(X) - g(\theta)\right)^2 \\
&= \mathbb{E}_\theta \left(\delta(X) - \mathbb{E}_\theta[\delta] + \mathbb{E}_\theta[\delta] - g(\theta)\right)^2 \\
&= \underbrace{\mathbb{E}_\theta \left(\delta(X) - \mathbb{E}_\theta[\delta]\right)^2}_{\mathrm{Var}_\theta(\delta)} + \underbrace{\mathbb{E}_\theta \left(\mathbb{E}_\theta[\delta] - g(\theta)\right)^2}_{\text{Bias}(\delta)^2} + \underbrace{2\mathbb{E}\left[(\delta - \mathbb{E}_\theta[\delta])(\mathbb{E}_\theta \delta - g(\theta))\right]}_{=0}
\end{aligned}$$

Logic: according to the bias-variance decomposition under squared error loss, if we restrict to unbiased estimators, comparing variance is equivalent to comparing risk

Def. UMVU

An unbiased estimator δ is uniformly minimum variance unbiased (UMVU) if

$$\mathrm{Var}_\theta(\delta) \leq \mathrm{Var}_\theta(\tilde{\delta}), \forall \theta \in \Omega$$

for any competing unbiased estimator $\tilde{\delta}$.

Does UMVU always exist?

No! Even unbiased estimators might not exist

Ex: estimate $\frac{1}{\theta^2}$ for X drawn from Uniform$(0, \theta)$

Does UMVU always exist?

No! Even unbiased estimators might not exist

Ex: estimate $\frac{1}{\theta^2}$ for X drawn from Uniform$(0, \theta)$

Def. U-estimable

We say $g(\theta)$ is U-estimable if there exists δ such that $\mathbb{E}_\theta \delta = g(\theta), \forall \theta \in \Omega$

Does UMVU exist under U-estimable assumption?

UMVU under U-estimable and given complete sufficient statistics

Theorem 4.4 in Keener, Lehmann-Scheffé

Suppose $T(X)$ is complete sufficient for $\mathscr{P} = \{P_\theta : \theta \in \Omega\}$. For any U-estimable $g(\theta)$, there is a unique (up to $\overset{\text{a.s.}}{=}$) UMVU estimator which is based on T.

Proof of Thm 4.4

- Existence
- Uniqueness
- UMVU

Extension to convex loss

Extension of Thm 4.4 to convex loss

Supppose $T(X)$ is complete sufficient for $\mathscr{P} = \{P_\theta : \theta \in \Omega\}$. Under a strictly convex loss, among all unbiased estimators, there is a unique (up to $\overset{\text{a.s.}}{=}$) uniformly minimum risk unbiased estimator which is based on T

Strategies for finding UMVU estimators

Two strategies for finding UMVU estimators:

- Directly find an unbiased estimator based on a complete sufficient T
- Find any unbiased estimator, then Rao-Blackwellize it.

Example 1

$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim}$ Poisson$(\theta), \theta > 0$.

- Find a UMVU estimator for θ
- Find a UMVU estimator for θ^2

Example 2

$X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta), \theta > 0.$

- Find a UMVU estimator for θ in two ways

In Example 2, is the UMVU estimator also a “good” (admissible) estimator in terms of total risk?

Summary

- Jensen's inequality for convex function. **Convex loss** allows us to rule out estimators with extra noise
- **Rao-Blackwell theorem** allows us to improve an estimator based on sufficient statistics T
- If unbiased estimator exists, complete sufficient statistics T exists, then **UMVU** estimator exists and is unique

What is next?

- Reflexion on the unbiasedness
- Information inequality

Thank you