Outline

9/14/23

1) Convex Loss

2) Rao-Blackwell Theorem

3) UMVU Estimators

4) Examples

Unbiased Estimation

Recall strategies to choose an estimator

1) Summarize risk by a scalar (avg or sup)

2) Restrict to a smaller class of estimators

Today: Unbiased estimation

estimand

Require $\mathbb{E}_\theta \delta(X) = g(\theta), \quad \forall \theta \in \Theta$

If we have complete sufficient stat $T(X)$,

- there is at most one unbiased $\delta^*(T(X))$

  (If $\mathbb{E}_\theta \delta_1(T) = \mathbb{E}_\theta \delta_2(T) = g(\theta) \quad \forall \theta$ then $\delta_1 \overset{a.s.}{=} \delta_2$)

- if it exists, it uniformly minimizes risk for any convex loss function

Convex Loss Functions

Recall $f(x)$ is **convex** if, for all x_1, x_2, all $\gamma \in [0,1]$

$$f(\gamma x_1 + (1-\gamma)x_2) \le \gamma f(x_1) + (1-\gamma) f(x_2)$$

strictly convex if $<$

Thm (Jensen) If f convex then

$$f(\mathbb{E}X) \le \mathbb{E} f(X) \quad \text{for any r.v. } X$$

f strictly convex then $<$ unless $X \overset{a.s.}{=} c$

Convex Loss $L(\theta, d)$ means convex **in d**

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

$$\begin{aligned} MSE(\theta; \delta) &= \mathbb{E}_\theta\left[(g(\theta) - \delta(X))^2\right] \\ &= \text{Bias}_\theta^2(\delta) + \text{Var}_\theta(\delta) \\ &= \text{Var}_\theta(T) \quad \text{if } \delta \text{ unbiased} \end{aligned}$$

Convex losses penalize us for making the estimator too noisy

Rao-Blackwell Theorem

Recipe to improve any $\delta(x)$ that violates the suff. principle.

Theorem (Rao-Blackwell)

Assume $T(X)$ sufficient, $\delta(X)$ estimator

Let $\bar{\delta}(T(X)) = \mathbb{E}\left[\delta(X) \mid T(X)\right]$

no θ

If $L(\theta, \delta)$ convex then $R(\theta; \bar{\delta}) \leq R(\theta; \delta)$

If strictly convex then $R(\theta; \bar{\delta}) < R(\theta; \delta)$

unless $\delta(X) \overset{a.s.}{=} \bar{\delta}(T(X))$ for all θ

Proof

$$
\begin{aligned}
R(\theta; \bar{\delta}) &= \mathbb{E}_\theta\left[L(\theta, \mathbb{E}[\delta \mid T])\right] \\
&\leq \mathbb{E}_\theta\, \mathbb{E}\left[L(\theta; \delta) \mid T\right] \\
&= R(\theta; \delta)
\end{aligned}
$$

$<$ if strictly, unless $\delta \overset{a.s.}{=} \bar{\delta}$ $\boxtimes$

$\bar{\delta}(T)$ called the Rao-Blackwellization of $\delta(X)$

U-estimable functions

Not all estimands have unbiased estimators:

Def We say $g(\theta)$ is U-estimable if $\exists\, \delta(X)$ with $\mathbb{E}_\theta \delta = g(\theta)\ \forall \theta$

Def $\delta(X)$ is uniform minimum variance unbiased (UMVU) if for any unbiased $\tilde{\delta}$,

$$\mathrm{Var}_\theta(\delta(X)) \le \mathrm{Var}_\theta(\tilde{\delta}(X)) \quad \forall \theta \in \Theta$$

Theorem For model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, assume:

i) $T(X)$ complete suff
ii) $g(\theta)$ U-estimable

Then there exists a unique estimator of the form $\delta^*(T(x))$, which

1) is UMVU and minimizes
2) uniformly minimizes risk among all unbiased estimators

Proof

"All Rao-Blackwellizations lead to δ^*"

Existence

Take any δ_0 unbiased for $g(\theta)$

Let $\delta^*(T) = \mathbb{E}[\delta_0 \mid T]$

no θ

$$\mathbb{E}_\theta \delta^* = \mathbb{E}_\theta\left[\mathbb{E}[\delta_0 \mid T]\right] = \mathbb{E}_\theta \delta_0 = g(\theta)$$

Uniqueness

If $\delta(T)$ unbiased then

$$\mathbb{E}_\theta\left[\delta^*(T) - \delta(T)\right] = 0 \quad \forall \theta \in \Theta$$

$$\Rightarrow \delta^*(T) \overset{a.s.}{=} \delta(T) \qquad \text{(Completeness)}$$

Optimality wrt any convex loss

Suppose $\delta(X)$ unbiased,

uniqueness

let $\bar{\delta}(T) = \mathbb{E}[X \mid T] \overset{a.s.}{=} \delta^*(T)$

Rao-Blackwell:

$$R(\theta; \delta^*) = R(\theta; \bar{\delta}) \leq R(\theta; \delta)$$

Hence, $MSE(\theta; \delta^*) \leq MSE(\theta; \delta)$

$$\begin{array}{ccc} MSE(\theta;\delta^*) & & MSE(\theta;\delta) \\ \shortparallel & & \shortparallel \\ Var_\theta(\delta^*) & & Var_\theta(\delta) \end{array}$$

So, δ^* UMVU $\boxtimes$

Finding the UMVUE

2 methods for finding UMVUE:

1) Find *any* unbiased estimator based on T
2) Find *any* unbiased estimator at all, then R-B'ize it.

<u>Ex</u>. $X_1, \dots, X_n \overset{iid}{\sim} \text{Pois}(\theta)$, $g(\theta) = \theta^2$

$$p_\theta^{(1)}(x) = \frac{\theta^x e^{-\theta}}{x!} \quad \theta > 0, \quad x = 0, 1, \dots$$

Complete suff. stat $T(X) = \sum X_i \sim \text{Pois}(n\theta)$

$$p_\theta^T(t) = \frac{(n\theta)^t e^{-n\theta}}{t!}$$

$\delta(T)$ unbiased

$$\Leftrightarrow \sum_{t=0}^{\infty} \delta(t) p_\theta^T(t) = \theta^2, \ \forall \theta$$

$$\Leftrightarrow \sum_{t=0}^{\infty} \delta(t) \frac{n^t}{t!} \theta^t = e^{n\theta} \theta^2 = \sum_{k=0}^{\infty} \frac{n^k}{k!} \theta^{k+2}, \ \forall \theta$$

Match terms in power series:

$$\delta(0) = \delta(1) = 0, \quad \delta(t) = \frac{n^{t-2}}{(t-2)!} \cdot \frac{t!}{n^t} \quad t \geq 2$$

$$\Rightarrow \delta(T) = \frac{T(T-1)}{n^2} \quad \left(\approx \left(\frac{T}{n}\right)^2 \text{ for large } T\right)$$

Alternatively, we could R-Bize $\delta_0(X) = X_1 X_2$

$$\mathbb{E}_\theta X_1 X_2 = (\mathbb{E}_\theta X_1)(\mathbb{E}_\theta X_2) = \theta^2$$

What is $\delta^*(T) = \mathbb{E}[X_1 X_2 \mid T]$?

$$X \mid T = t \sim \text{Multinom}(n, \tfrac{1}{n} 1_n)$$

$$X_1 \mid T = t \sim \text{Binom}(n, \tfrac{1}{n})$$

$$\Rightarrow \mathbb{E}[X_1 \mid T] = T/n$$

$$\text{Var}(X_1 \mid T) = T \cdot \tfrac{1}{n}(1 - \tfrac{1}{n}) = \frac{T(n-1)}{n^2}$$

$$\mathbb{E}[X_2 \mid T, X_1] = \frac{T - X_1}{n-1}$$

$$\begin{aligned}
\mathbb{E}[X_1 X_2 \mid T] &= \mathbb{E}[X_1 \mathbb{E}[X_2 \mid T, X_1] \mid T] \\
&= \mathbb{E}\left[\frac{T}{n-1} X_1 - \frac{1}{n-1} X_1^2 \mid T\right] \\
&= \frac{T^2}{n(n-1)} - \frac{1}{n-1}\left(\frac{T^2}{n^2} + \frac{T(n-1)}{n^2}\right) \\
&= \frac{1}{n^2(n-1)}\left(T^2 n - T^2 - T(n-1)\right) \\
&= \frac{T(T-1)}{n^2}
\end{aligned}$$

Ex $X_1, \dots, X_n \overset{iid}{\sim} U[0, \theta] \qquad \theta > 0$

$T = X_{(n)}$ complete suff.

$$p_\theta^T = \frac{n}{\theta^n} t^{n-1} \, 1\{t \le \theta\}$$

$$\mathbb{E}_\theta T = \int_0^\theta t \frac{n}{\theta^n} t^{n-1} dt = \frac{n}{n+1} \theta$$

$$\Rightarrow \frac{n+1}{n} T \text{ is UMVU}$$

Alternate $2X_1$ is unbiased

$$X_1 \mid T \sim \begin{cases} T & \text{wp } \frac{1}{n} \\ U[0, T] & \text{wp } \frac{n-1}{n} \end{cases}$$

$$\Rightarrow \mathbb{E}[2X_1 \mid T] = 2T \cdot \frac{1}{n} + T \cdot \frac{n-1}{n}$$

$$= \frac{n+1}{n} T$$

Actually, $\frac{n+1}{n} T$ is inadmissible too!

Keener shows $\frac{n+2}{n+1} T$ has best MSE for any estimator $c \cdot T$.

Raises question: why do we require 0 bias?

Doubts about unbiasedness

The UMVUE might be very inefficient, or inadmissible, or just dumb, in cases where another approach makes much more sense

Ex. $X \sim \text{Bin}(1000, \theta)$

Estimate $g(\theta) = \mathbb{P}_\theta(X \geq 500)$

UMVUE is $1\{X \geq 500\}$ (why?)

$\Rightarrow$ $X = 500$? Conclude $g(\theta) = 100\%$

$X = 499$? Conclude $g(\theta) = 0\%$

This is not epistemically reasonable!!

Could do much better with e.g. MLE or a Bayes estimator.

In fact, our theorem should make us suspicious of UMVUE's: every idiotic function of T is a UMVUE (of its own expectation)

Gaussian Sequence Model

$X_i \overset{iid}{\sim} N(\mu_i, 1) \qquad i = 1, \dots, d \qquad$ indep.

or $\quad X \sim N_d(\mu, I_d) \qquad \mu \in \mathbb{R}^d$, estimate $\rho^2 = \|\mu\|^2$

X is complete sufficient

$$\begin{aligned} \mathbb{E}_\mu \|X\|^2 &= \mathbb{E}_0\left[\|\mu + X\|^2\right] \\ &= \|\mu\|^2 + \mathbb{E}_0\|X\|^2 + \cancelto{0}{2\mathbb{E}_0[\mu' X]} \\ &= \|\mu\|^2 + d \end{aligned}$$

$$\Rightarrow \quad \delta(x) = \|X\|^2 - d$$

If $\mu = 0$, $\quad \delta(x) < 0 \quad$ about half the time!

$$(\|X\|^2 - d)_+ = \max(0, \|X\|^2 - d)$$

strictly dominates UMVU