Outline

1) Convex Loss

2) Rao - Blackwell Theorem

3) UMVU Estimators

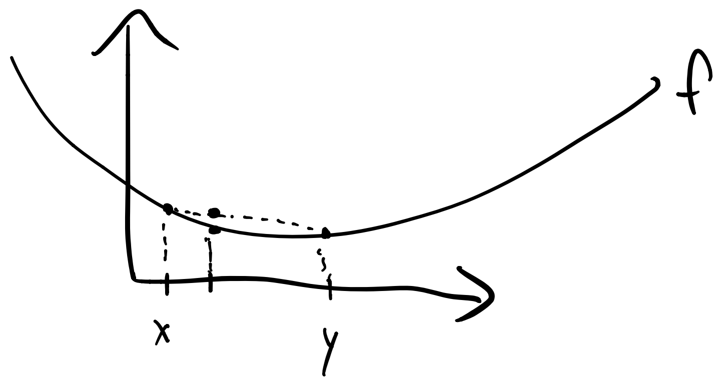
4) Examples

Convex Loss Functions

Def $f(x)$ is convex if, $\gamma \in (0,1)$

$$f(\gamma x + (1-\gamma)y) \leq \gamma f(x) + (1-\gamma) f(y)$$

strictly convex if $<$



Thm (Jensen) If f convex then

$$f(\mathbb{E}X) \leq \mathbb{E}f(X)$$ for any prob. measure $\mathbb{P}$

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

X can be a random vector

If $L(\theta, d)$ is convex in d then it penalizes us for adding extra noise to an estimate

Let $\tilde{\delta}(X) = \delta(X) + \varepsilon$,

where ε is mean-zero noise $(\varepsilon \perp\!\!\!\perp X)$

$$R(\theta; \tilde{\delta}) = \mathbb{E}_\theta\left[\mathbb{E}\left[L(\theta, \delta(X) + \varepsilon) \mid X\right]\right]$$

$$\geq \mathbb{E}_\theta\left[L(\theta, \delta(X))\right]$$

$$= R(\theta; \delta)$$

$>$ if L strictly convex, $\varepsilon \overset{a.s.}{\neq} 0$

Rao-Blackwell Theorem

For estimation with convex loss there is a very clear reason to adhere to the sufficiency 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 $L(\theta, d)$ 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 $\delta \overset{a.s.}{=} \bar{\delta}$ $\boxtimes$

$\delta \rightsquigarrow \bar{\delta}$ is called Rao-Blackwellization.

Bias-Variance Decomposition

The **bias** of $\delta(X)$ is $\mathbb{E}_\theta[\delta(X) - g(\theta)]$

We say $\delta(X)$ is **unbiased** if

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

Bias-Variance Decomposition makes explicit the (MSE) price we pay for noise

$$\begin{aligned}
\mathrm{MSE}(\theta; \delta) &= \mathbb{E}_\theta\left[(\delta(X) - g(\theta))^2\right] \\
&= \mathbb{E}_\theta\Big[\big(\underbrace{\delta - \mathbb{E}_\theta\delta}_{\text{mean-zero}} + \underbrace{\mathbb{E}_\theta\delta - g(\theta)}_{\text{deterministic}}\big)^2\Big] \\
&= \left(\mathbb{E}_\theta\delta - g(\theta)\right)^2 + \mathbb{E}_\theta\left[(\delta - \mathbb{E}_\theta\delta)^2\right] \\
&\quad + \cancel{2\,\mathbb{E}_\theta\left[(\delta - \mathbb{E}_\theta\delta)(\mathbb{E}_\theta\delta - g(\theta))\right]} \to 0 \\
&= \mathrm{Bias}_\theta(\delta)^2 + \mathrm{Var}_\theta(\delta)
\end{aligned}$$

Next topic is **unbiased estimation**:

[what if we require bias $= 0$ and do as well as we can on variance?]

An unbiased estimator doesn't always exist.

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

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

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

Theorem 4.4 Suppose $T(X)$ is complete sufficient for $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Then, for any U-estimable $g(\theta)$, there is a unique (up to $\overset{a.s.}{=}$) UMVU estimator of the form $\delta(T(X))$

Proof Assume δ_0 is unbiased for $g(\theta)$

Existence

Then $\delta(X) = \underset{\text{no } \theta}{\mathbb{E}}[\delta_0 \mid T]$ is unbiased:

$$\mathbb{E}_\theta \delta = \mathbb{E}_\theta\big[\mathbb{E}[\delta_0 \mid T]\big] = \mathbb{E}_\theta \delta_0 = g(\theta)$$

Uniqueness

If $\tilde{\delta}(T)$ unbiased then

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

$$\implies \delta(T) \overset{a.s.}{=} \tilde{\delta}(T) \qquad \text{(Completeness)}$$

UMVU

<u>Rao - Blackwell</u>: Suppose $\delta^*(X)$ unbiased

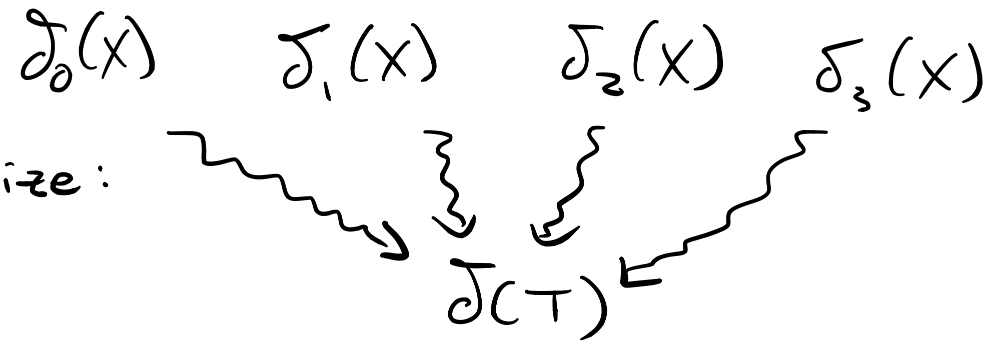
Then $\mathrm{Var}_\theta(\delta^*(X)) \geq \mathrm{Var}_\theta\left[\mathbb{E}\left[\delta^*(X) \mid T\right]\right]$

$$= \mathrm{Var}_\theta\, \delta$$

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

In a picture:

R-B'ize:



R-Bizing improves the estimator, and all R-Bizations are $\overset{a.s.}{=}$ by completeness.

<u>Note</u>: same proof works for any convex loss: $\delta(T(X))$ has uniformly minimum risk for <u>any</u> convex L

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.

Ex. $X_1, \dots, X_n \overset{iid}{\sim} \text{Pois}(\theta) \qquad \theta > 0$

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

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

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

Estimate $g(\theta) = \theta^2$

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

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

Match terms in power series:

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

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

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

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

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

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

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

<u>Alternate</u> $2X_1$ 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 $c \cdot T$.

<u>Raises question</u>: why do we require 0 bias?