# 15. Minimum Variance Unbiased Estimation

**ECE 830, Spring 2014**

# Bias-Variance Trade-Off

Recall that

$$\mathsf{MSE}(\widehat{\theta}) = \mathsf{Bias}^2(\widehat{\theta}) + \mathsf{Var}(\widehat{\theta}).$$

In general, the minimum MSE estimator has non-zero bias and non-zero variance.

We can reduce bias only at a potential increase in variance.

Conversely, modifying the estimator to reduce the variance may lead to an increase in bias.

## Example:

Let

$$x_n = A + w_n$$
$$w_n \sim \mathcal{N}\left(0, \sigma^2\right)$$
$$\widetilde{A} = \frac{\alpha}{N} \sum_{n=1}^{N} x_n$$

where $\alpha$ is an arbitrary constant. If

$$S_N \equiv \frac{1}{N} \sum_{n=1}^{N} x_n,$$

then

$$\widetilde{A} \quad =$$
$$S_N \quad \sim$$

## Example: (cont.)

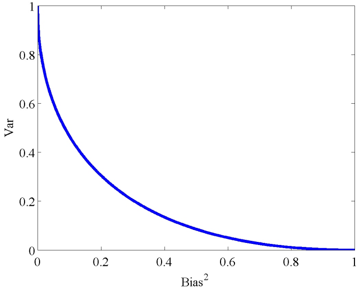
Let's find the value of $\alpha$ that minimizes the MSE.

$$\text{Var}\left(\widetilde{A}\right) = \text{Var}\left(\alpha S_N\right) = \quad \text{Var}\left(S_N\right) =$$

$$\text{Bias}\left(\widetilde{A}\right) = \mathbb{E}\left[\widetilde{A}\right] - A = \mathbb{E}\left[\ \ S_N\right] - A = \qquad - A =$$

Thus the MSE is

$$\text{MSE}\left(\widetilde{A}\right) =$$



Aside: alternatively, we could have computed the MSE as follows

$$
\begin{aligned}
\mathbb{E}[x_i x_j] &= \begin{cases} A^2 + \sigma^2 & , i = j \\ A^2 & , i \neq j \end{cases} \\
\mathsf{MSE}\left(\widetilde{A}\right) &= \mathbb{E}\left[\left(\widetilde{A} - A\right)^2\right] \\
&= \mathbb{E}\left[\widetilde{A}^2\right] - 2\mathbb{E}\left[\widetilde{A}\right] A + A^2 \\
&= \alpha^2 \mathbb{E}\left[\frac{1}{N^2} \sum_{i,j=1}^{N} x_i x_j\right] - 2\alpha \mathbb{E}\left[\frac{1}{N} \sum_{n=1}^{N} x_n\right] A + A^2 \\
&= \alpha^2 \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}[x_i x_j] - 2\alpha \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[x_n] A + A^2 \\
&= \alpha^2 \left(A^2 + \frac{\sigma^2}{N}\right) - 2\alpha A^2 + A^2 \\
&= \underbrace{\frac{\alpha^2 \sigma^2}{N}}_{\mathsf{Var}(\widetilde{A})} + \underbrace{(\alpha - 1)^2 A^2}_{\mathsf{Bias}^2(\widetilde{A})}
\end{aligned}
$$

## So how practical is the MSE as a *design criterion*?

In the previous example, the MSE is minimized when

$$\frac{d\text{MSE}\left(\widetilde{A}\right)}{d\alpha} =$$

$$\Rightarrow \alpha^* =$$

The optimal (in an MSE sense) value $\alpha^*$ depends on the unknown parameter $A$! Therefore, the estimator is not realizable. This phenomenon occurs for many classes of problems.

**We need an alternative to direct MSE minimization.**

Note that in the above example, the problematic dependence on the parameter $(A)$ enters through the Bias component of the MSE. This occurs in many situations. Thus a reasonable alternative is to

constrain the estimator to be unbiased, and then find the estimator that produces the minimum variance (and hence provides the minimum MSE among all unbiased estimators).

Note: Sometimes no unbiased estimator exists and we cannot proceed at all in this direction.

**Definition: Minimum Variance Unbiased Estimator**

$\widehat{\theta}$ is a *minimum variance unbiased estimator* (MVUE) for $\theta$ if

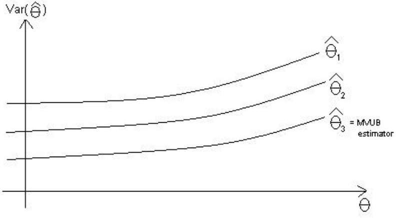
1. $\mathbb{E}\widehat{\theta} = \theta \;\forall\; \theta \in \Theta$
2. If $\mathbb{E}\widehat{\theta_0} = \theta \;\forall\; \theta \in \Theta$, then $\text{Var}\left(\widehat{\theta}\right) \leq \text{Var}\left(\widehat{\theta_0}\right) \;\forall\; \theta \in \Theta$.

# Existence of the Minimum Variance Unbiased Estimator (MVUE)

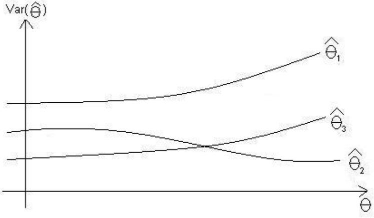
Does an MVUE estimator exist? Suppose there exist three unbiased estimators:

$$\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3$$

Two possibilities exist.




$\widehat{\theta}_3$ is MVUE




no MVUE exists!

## Example:

Suppose we observe a single scalar realization $x$ of

$$X \sim \mathsf{Unif}\left(0, 1/\theta\right), \theta > 0.$$

An unbiased estimator of $\theta$ does not exist. To see this, note that

$$p\left(x|\theta\right) = \theta \cdot I_{[0,1/\theta]}\left(x\right).$$

If $\widehat{\theta}$ is unbiased, then

$$\forall \theta > 0,\ \ \theta = \mathbb{E}\left[\widehat{\theta}\right] =$$

$$\Longrightarrow$$

$$\Longrightarrow$$

But if this is true for all $\theta$, then we have $\widehat{\theta}\left(x\right) = 0$, which is not an unbiased estimator.

# Finding the MVUE Estimator

There is no simple, general procedure for finding the MVUE estimator. In the next several lectures we will discuss several approaches:

1. Find a sufficient statistic and apply the Rao-Blackwell theorem
2. Determine the so-called Cramer-Rao Lower Bound (CRLB) and verify that the estimator achieves it.
3. Further restrict the estimator to a class of estimators (e.g., linear or polynomial functions of the data)

# Recipe for finding a MVUE

**(1)** Find a *complete* sufficient statistic $t = T(X)$.

**(2)** Find any *unbiased estimator* $\widehat{\theta}_0$ and set

$$\widehat{\theta}(X) := \mathbb{E}[\widehat{\theta}_0(X) | t = T(X)]$$

*or* find a function $g$ such that

$$\widehat{\theta}(X) = g(T(X))$$

is unbiased.

These notes answer the following questions:

1. What is a sufficient statistic?
2. What is a complete sufficient statistic?
3. What does step (2) do above?
4. Is this estimator unique?
5. How do we know it's the MVUE?

**Definition: Sufficient statistic**

Let $X$ be an $N$-dimensional random vector and let $\theta$ denote a $p$-dimensional parameter of the distribution of $X$. The statistic $t := T(X)$ is a *sufficient statistic* for $\theta$ if and only if the conditional distribution of $X$ given $T(X)$ is independent of $\theta$.

See lecture 4 for more information on Sufficient Statistics and how to find them.

# Minimal and Complete Sufficient Statistics

**Definition: Minimal Sufficient Statistic**

A sufficient statistic $t$ is said to be *minimal* if the dimension of $t$ cannot be reduced and still be sufficient.

**Definition: Complete sufficient statistic**

A sufficient statistic $t := T(X)$ is *complete* if for all real-valued functions $\phi$ which satisfy

$$(\mathbb{E}[\phi(t)|\theta] = 0 \forall \theta)$$

we have

$$(\mathbb{P}[\phi(t) = 0|\theta] = 1 \forall \theta)$$

Under very general conditions, if $t$ is a *complete* sufficient statistic, then $t$ is *minimal*.

## Example: Bernoulli trials

Consider $N$ independent Bernoulli trials

$$x_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta), \theta \in [0,1].$$

Recall $k = \sum_{n=1}^{N} x_i$ is sufficient for $\theta$. Now suppose $\mathbb{E}[\phi(k)|\theta] = 0$ for all $\theta$. But

$$\begin{aligned} \mathbb{E}[\phi(k)|\theta] &= \\ &= \end{aligned}$$

where poly$(\theta)$ is an $N^{\text{th}}$ degree polynomial. Then

$$\begin{aligned} &\text{poly}(\theta) = 0 \forall \theta \in [0,1] \\ \implies &\ \text{poly}(\theta) \text{ is the zero polynomial} \\ \implies &\ \phi(k) \\ \implies & \end{aligned}$$

# Rao-Blackwell Theorem

**Rao-Blackwell Theorem**

Let $\boldsymbol{Y}, \boldsymbol{Z}$ be random variables and define the function

$$g(\boldsymbol{z}) := \mathbb{E}[\boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{z}].$$

Then

$$\mathbb{E}[g(\boldsymbol{Z})] = \mathbb{E}[\boldsymbol{Y}]$$

and

$$\mathsf{Var}(g(\boldsymbol{Z})) \leq \mathsf{Var}(\boldsymbol{Y})$$

with equality iff $\boldsymbol{Y} = g(\boldsymbol{Z})$ almost surely.

Note that this version of Rao-Blackwell is quite general and has nothing to do with estimation of parameters. However, we can apply it to parameter estimation as follows.

Consider $X \sim p(x|\theta)$. Let $\widehat{\theta}_1$ be an unbiased estimator of $\theta$ and let $t = T(x)$ be a sufficient statistic for $\theta$. Apply Rao-Blackwell with

$$\begin{aligned} \boldsymbol{Y} &:= \widehat{\theta}_1(x) \\ \boldsymbol{Z} &:= t = T(x). \end{aligned}$$

Consider the new estimator

$$\widehat{\theta}_2(x) = g(T(x)) = \mathbb{E}[\widehat{\theta}_1(X)|T(X) = t].$$

Then we may conclude:

1. $\widehat{\theta}_2$ is unbiased
2. $\mathrm{Var}(\widehat{\theta}_2) \leq \mathrm{Var}(\widehat{\theta}_1)$

In words, if $\widehat{\theta}_1$ is any unbiased estimator, then smoothing $\widehat{\theta}_1$ with respect to a sufficient statistic decreases the variance while preserving unbiasedness.

Therefore, we can restrict our search for the MVUE to functions of a sufficient statistic.

# The Rao-Blackwell Theorem

### Rao-Blackwell Theorem, special case

Let $X$ be a random variable with pdf $p(X|\theta)$ and let $t(X)$ be a sufficient statistic. Let $\widehat{\theta}_1(x)$ be an estimator of $\theta$ and define

$$\widehat{\theta}_2(t) := \mathbb{E}\left[\widehat{\theta}_1(X)|t(X)\right].$$

Then

$$\mathbb{E}[\widehat{\theta}_2(T)] = \mathbb{E}[\widehat{\theta}_1(X)]$$

and

$$\text{Var}(\widehat{\theta}_2(T)) \leq \text{Var}(\widehat{\theta}_1(X))$$

with equality iff $\widehat{\theta}_1(X) \equiv \widehat{\theta}_2(t(X))$ with probability one (almost surely).

# Rao-Blackwell Theorem in Action

Suppose we observe 2 independent realizations from a $\mathcal{N}(\mu, \sigma^2)$ distribution. Denote these observations $x_1$ and $x_2$, with $X = [x_1, x_2]^T$. Consider the simple estimator of $\mu$:

$$\begin{aligned} \hat{\mu} =& x_1 \\ \mathbb{E}[\widehat{\mu}] =& \\ \text{Var}[\widehat{\mu}] =& \end{aligned}$$

The MSE is therefore:

Intuitively, we expect that the sample mean should be a better estimator since

$$\widetilde{\mu} = \frac{1}{2}(x_1 + x_2)$$

averages the two observations together.

# Is this the best possible estimator?

Let's find a sufficient statistic for $\mu$:

$$\begin{aligned} p(x_1, x_2) &= \frac{1}{2\pi\sigma^2} e^{-(x_1-\mu)^2/2\sigma^2} e^{-(x_2-\mu)^2/2\sigma^2} \\ &= \\ &= \end{aligned}$$

The Rao-Blackwell Theorem states that:

$$\mu^* = \mathbb{E}[\widehat{\mu}|t]$$

is as good as or better than $\widehat{\mu}$ in terms of estimator variance. (See Scharf p94.) What is $\mu^*$? First we need to compute the mean of the conditional density $p(\widehat{\mu}|t)$ or $p(x_1|t)$

$$\begin{aligned} p(x_1|t) &= \frac{p(x_1, t)}{p(t)} \\ p(x_1, t) &= \\ p(t) &= \\ \\ \mathbb{E}(t) &= \\ \mathsf{Var}(t) &= \end{aligned}$$

$$
\begin{aligned}
p(x_1|t) &= \frac{\frac{1}{2\pi\sigma^2}}{\frac{1}{\sqrt{4\pi\sigma^2}}} \exp\left[\frac{-1}{2\sigma^2}\left((x_1-\mu)^2 + (t-x_1-\mu)^2 - (t-2\mu)^2/2\right)\right] \\
&= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[\begin{array}{r} \frac{-1}{2\sigma^2}\left(x_1^2 - 2\mu x_1 + \mu^2 + t^2 - 2x_1 t + x_1^2 - 2\mu t + \right. \\ \left. +2\mu x_1 + \mu^2 - t^2/2 + 4\mu t/2 - 4\mu^2/2\right) \end{array}\right] \\
&= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2}\left(2x_1^2 - 2x_1 t + t^2/2\right)\right] \\
&= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[\frac{-(x_1 - t/2)^2}{\sigma^2}\right] \\
\Rightarrow x_1|t &\sim
\end{aligned}
$$

$$
\begin{aligned}
\mu^* =& \mathbb{E}[\widehat{\mu}|t] = \\
\text{Var}(\mu^*) &= \\
\Rightarrow \text{MSE}(\mu^*) &=
\end{aligned}
$$

# The Lehmann-Scheffe Theorem

The Rao-Blackwell Theorem tells us how to decrease the variance of an unbiased estimator. But when can we know that we get a MVUE?

Answer: When $t$ is a complete sufficient statistic.

**Lehmann-Scheffe Theorem**

If $t$ is *complete*, there is at most *one* unbiased estimator that is a function of $t$.

# Proof

Suppose

$$\mathbb{E}[\widehat{\theta}_1] = \mathbb{E}[\widehat{\theta}_2] = \theta$$
$$\widehat{\theta}_1(X) := g_1(T(X))$$
$$\widehat{\theta}_2(X) := g_2(T(X)).$$

Define

$$\phi(t) := g_1(t) - g_2(t).$$

Then

$$\mathbb{E}[\phi(t)] =$$

By definition of completeness, we have

In other words

$$\widehat{\theta}_1 = \widehat{\theta}_2 \text{ with probability } 1.$$

■

# Recipe for finding a MVUE

This result suggests the following method for finding a MVUE:

**(1)** Find a complete sufficient statistic $t = T(X)$.

**(2)** Find any unbiased estimator $\widehat{\theta}_0$ and set

$$\widehat{\theta}(X) := \mathbb{E}[\widehat{\theta}_0(X)|t = T(X)]$$

or find a function $g$ such that

$$\widehat{\theta}(X) = g(T(X))$$

is unbiased.

# Rao-Blackwell and Complete Suff. Stats.

**Theorem**

If $\widehat{\theta}$ is constructed by the recipe above, then $\widehat{\theta}$ is the *unique* MVUE.

**Proof:** Note that in either construction, $\widehat{\theta}$ is a function of $t$. Let $\widehat{\theta}_1$ be any unbiased estimator. We must show that

$$\mathsf{Var}(\widehat{\theta}) \leq \mathsf{Var}(\widehat{\theta}_1).$$

Define

$$\widehat{\theta}_2(X) := \mathbb{E}[\widehat{\theta}_1(X)|t = T(X)].$$

By Rao-Blackwell, it suffices to show

$$\mathsf{Var}(\widehat{\theta}) \leq \mathsf{Var}(\widehat{\theta}_2).$$

# Proof (cont.)

But $\widehat{\theta}$ and $\widehat{\theta}_2$ are both unbiased and functions of a complete sufficient statistic

To show uniqueness, in the above argument suppose $\text{Var}(\widehat{\theta}_1) = \text{Var}(\widehat{\theta})$. Then the Rao-Blackwell bound holds with equality

## Example: Uniform distribution.

Suppose $X = [x_1 \cdots x_N]^T$ where

$$x_i \overset{\text{iid}}{\sim} \text{Unif}[0, \theta], i = 1, \ldots, N.$$

What is an unbiased estimator of $\theta$?

$$\widehat{\theta}_1 = \frac{2}{N} \sum_{i=1}^{N} x_i$$

is unbiased. However, it is not MVUE.

## Example: (cont.)

From the Fisher-Neyman factorization theorem,

$$\begin{aligned} p(X|\theta) &= \prod_{i=1}^{N} \frac{1}{\theta} I_{[0,\theta]}(x_i) \\ &= \underbrace{\frac{1}{\theta^N} I_{[\max_i x_i, \infty)}(\theta)}_{b_\theta(t)} \cdot \underbrace{I_{(-\infty, \min_i x_i]}(0)}_{a(X)} \end{aligned}$$

we see that

$$T = \max_i x_i$$

is a sufficient statistic. It is left as an exercize to show that $T$ is in fact complete. Since $\widehat{\theta}_1$ is not a function of $T$, it is not MVUE. However,

$$\widehat{\theta}_2(X) = \mathbb{E}[\widehat{\theta}_1(X)|t = T(X)]$$

is the MVUE.