ECE531 Lecture 8: Non-Random Parameter Estimation

D. Richard Brown III

Worcester Polytechnic Institute

23-March-2011

Introduction

- Recall the two basic classes of estimation problems:
  - Known prior $\pi(\theta)$: Bayesian estimation.
  - Unknown prior (θ is not a RV): Non-random parameter estimation.
- In non-random parameter estimation problems, we can still compute the risk of estimator $\hat{\theta}(y)$ when the true parameter is θ:

$$R_\theta(\hat{\theta}) := \mathrm{E}_\theta\left[C_\theta(\hat{\theta}(Y))\right] = \int_{\mathcal{Y}} C_\theta(\hat{\theta}(y))p_Y(y\,;\,\theta)\,dy$$

  where E_θ means the expectation parameterized by θ and $C_\theta(\hat{\theta}) : \Lambda \times \Lambda \mapsto \mathbb{R}$ is the cost of the parameter estimate $\hat{\theta} \in \Lambda$ given the true parameter $\theta \in \Lambda$.
- We cannot, however, compute any sort of average risk

$$r(\hat{\theta}) = \mathrm{E}[R_\Theta(\hat{\theta})]$$

  since we have no distribution on the random parameter Θ.

MMSE Criterion for Non-Random Parameter Estimation?

Recall, when the unknown parameter was random $\Theta \sim \pi(\theta)$, we could compute the mean squared error of an estimator $\hat{\theta}(y)$ as

$$\text{MSE}(\hat{\theta}) \;=\; \mathrm{E}\left[\|\Theta - \hat{\theta}(Y)\|_2^2\right] \qquad \text{(random parameter)}$$

where the expectation is evaluated with respect to the joint pdf $p_{Y,\Theta}(y,\theta)$.

This concept of MSE can be extended to non-random parameter estimation by fixing the parameter θ and taking the expectation with respect to only the observations, i.e.

$$\text{MSE}(\hat{\theta},\theta) \;=\; \mathrm{E}\left[\|\theta - \hat{\theta}(Y)\|_2^2\right] \qquad \text{(non-random parameter)}$$

where the expectation is evaluated with respect to $p_Y(y\,;\;\theta)$.

Question: Is it possible to find a non-random parameter estimator that minimizes the MSE for all θ?

Non-Random Parameter MMSE Example

Suppose you receive one observation

$$Y = \theta + W$$

where $\theta \in \mathbb{R}$ and $W \sim \mathcal{N}(0,1)$. Suppose further that your estimator is

$$\hat{\theta}(y) = ay$$

where a is a scalar parameter that you will specify to minimize the MSE. The non-random parameter MSE is then

$$\begin{aligned} \text{MSE}(\hat{\theta},\theta) &= \mathrm{E}\left[(\theta - aY)^2\right] \\ &= \theta^2 - 2a\theta\mathrm{E}[Y] + a^2\mathrm{E}[Y^2] \\ &= \theta^2 - 2a\theta^2 + a^2(\theta^2+1) \end{aligned}$$

Take a derivative with respect to a, set it to zero, and solve for a to get

$$a = \frac{\theta^2}{\theta^2+1}.$$

Hence $\hat{\theta}_{\text{mmse}}(y) = \frac{\theta^2}{\theta^2+1}y$. What is the problem with this result?

Our Approach: Consider Only Unbiased Estimators

Since MMSE and UMP non-random parameter estimators are unlikely to exist in most cases, we will consider only the class of **unbiased estimators**.

Definition

An estimator $\hat{\theta}(y)$ is unbiased if

$$\mathrm{E}_\theta\left[\hat{\theta}(Y)\right] \;=\; \theta$$

for all $\theta \in \Lambda$.

Remarks:

- This class excludes trivial estimators like $\hat{\theta}(y) \equiv \theta_0$ (constant).
- Under the squared-error cost assignment, the "parameterized risk" (conditional risk) of estimators in this class

$$R_\theta(\hat{\theta}) = \mathrm{E}_\theta\left[\|\theta - \hat{\theta}(Y)\|_2^2\right] = \sum_i \mathrm{E}_{\theta_i}\left[(\hat{\theta}_i(Y) - \theta_i)^2\right] = \sum_i \mathrm{var}_{\theta_i}\left[\hat{\theta}_i(Y)\right]$$

- The goal: **find an unbiased estimator with minimum variance**.

Minimum Variance Unbiased Estimators

Definition

A minimum-variance unbiased estimator $\hat{\theta}_{\mathsf{mvu}}(y)$ is an unbiased estimator satisfying

$$\hat{\theta}_{\mathsf{mvu}}(y) = \arg \min_{\hat{\theta}(y) \in \Omega} R_\theta(\hat{\theta}(y))$$

for all $\theta \in \Lambda$ where Ω is the set of all unbiased estimators.

Remarks:

- Finding an MVU estimator is still a multi-objective optimization problem. You have to find one estimator to minimize the variance at all $\theta \in \Lambda$.
- The estimator can not be a function of θ.
- MVU estimators do not always exist (see Example 2.3 in Kay I).
- We will see, however, that lots of problems do yield MVU estimators.

Example: Estimating a Constant in White Gaussian Noise

Suppose we have random observations given by

$$Y_k = \theta + W_k \qquad k = 0, \dots, n-1$$

where $W_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ with $\theta \in \mathbb{R}$.

Suppose we go with an estimator that performs the sample mean:

$$\hat{\theta}(y) = \frac{1}{n} \sum_{k=0}^{n-1} y_k$$

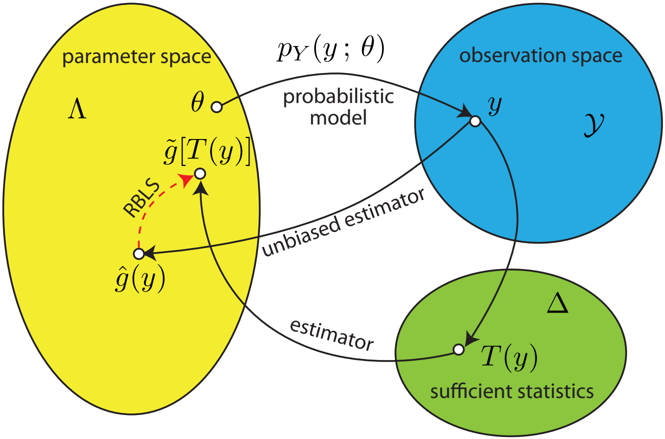
- Is this estimator unbiased? Yes (easy to check).
- Is this estimator MVU? The variance of the estimator can be calculated as $\text{var}_\theta \left[\hat{\theta}(Y) \right] = \frac{\sigma^2}{n}$. But answering the question as to whether this estimator is MVU or not will require more work.

Finding MVU Estimators

There is no "plug-and-chug" method that you can always follow to find an MVU estimator. We will cover two common approaches that work in many cases:

1. The Rao-Blackwell-Lehmann-Sheffe (RBLS) theorem (Kay I: Chapter 5)
   - Finding a "complete sufficient statistic" for the observations.
   - Performing a conditional expectation to get the MVU estimator.
2. The Cramer-Rao lower bound (Kay I: Chapters 3-4)
   - Guess at a good estimator and check if the variance achieves the theoretical minimum (CRLB).

A Generalized Model for Non-Random Parameter Estimation Problems



RBLS: A Procedure for Finding MVU Estimators

To find an MVU estimator for the non-random parameter θ, we can follow a three-step procedure:

1. Find a **complete sufficient statistic** $T : \mathcal{Y} \to \Delta$ for the family of pdfs $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$ parameterized by θ.
2. Find *any* unbiased estimator $\hat{g}(y)$ of θ.
3. Compute $\tilde{g}[T(y)] = \mathrm{E}_\theta[\hat{g}(Y)\,|\,T(Y) = T(y)]$.

The **Rao-Blackwell-Lehmann-Sheffe Theorem** says that $\tilde{g}[T(y)]$ will be a MVU estimator of the non-random parameter θ.

Some Intuition About Sufficient Statistics

Suppose we have observations given by

$$Y_k = \theta + W_k \qquad k = 0, \dots, n-1$$

where $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. It turns out that

$$\hat{\theta}(y) = \frac{1}{n} \sum_{k=0}^{n-1} y_k$$

is a MVU estimator (we will show this soon).

- The set of observations is $\mathcal{S} = \{y_0, \dots, y_{n-1}\}$.
- If we threw away some of the observations, would we still be able to compute the MVU estimator?
- What about these observation sets?

$$\mathcal{S}' = \{y_0\}$$
$$\mathcal{S}'' = \{y_0 + y_1, y_2, \dots, y_{n-1}\}$$
$$\mathcal{S}''' = \{y_0 + y_1 + \cdots + y_{n-1}\}$$

Some Intuition About Sufficient Statistics

- The original set of observations is always a "sufficient statistic". But there are often smaller sets that contain the relevant information.
- A sufficient statistic **summarizes** all of the information in the original set of observations so that, if you condition the pdf of the observations on the sufficient statistic, the pdf no longer depends on the unknown parameter.
- After observing a sufficient statistic, the original set of observations may be discarded. They add no additional information.
- There may be many different sufficient statistics.
- The sufficient statistic is not allowed to be a function of the unknown parameter θ. It may only be a function of the observations (and the number of observations).

Verifying a Sufficient Statistic (part 1)

Suppose we have observations given by

$$Y_k = \theta + W_k \qquad k = 0, \dots, n-1$$

where $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ and you want to verify

$$T(y) = \sum_{k=0}^{n-1} y_k$$

is a sufficient statistic. Let $Z = T(Y)$. Clearly Z is a random variable, with pdf parameterized by θ. What is the pdf of Z?

Verifying a Sufficient Statistic (part 2)

To verify the pdf of Y no longer depends on the unknown parameter when we observe $Z = T(Y) = t$, we need to compute

$$p_Y(y \,|\, Z = t\,;\, \theta) = \frac{p_{Y,Z}(y, t\,;\, \theta)}{p_Z(t\,;\, \theta)}.$$

Since $Z \sim \mathcal{N}(n\theta, n\sigma^2)$, the denominator of the RHS is easy:

$$p_Z(t\,;\, \theta) = \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left\{\frac{-(t - n\theta)^2}{2n\sigma^2}\right\}$$

The joint distribution $p_{Y,Z}(y, t\,;\, \theta)$ is a little bit tricky. The random variables Y and Z are **functionally dependent** since $Z = T(Y)$. This means that the joint pdf is zero everywhere unless the dummy variables satisfy $t = \sum_{k=0}^{n-1} y_k$.

Verifying a Sufficient Statistic (part 3)

Hence, the numerator of the RHS is

$$p_{Y,Z}(y,t\,;\,\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta)^2\right\}\delta\left(\sum_{k=0}^{n-1} y_k - t\right)$$

Skipping the algebraic details, we can compute the conditional pdf as

$$\begin{aligned} p_Y(y\,|\,Z=t\,;\,\theta) &= \frac{\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{\frac{-1}{2\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta)^2\right\}\delta\left(\sum_{k=0}^{n-1}y_k - t\right)}{\frac{1}{\sqrt{2\pi n\sigma^2}}\exp\left\{\frac{-(t-n\theta)^2}{2n\sigma^2}\right\}} \\ &= \frac{\sqrt{n}}{(2\pi\sigma^2)^{\frac{n-1}{2}}}\exp\left\{\frac{-1}{2\sigma^2}\sum_{k=0}^{n-1}y_k^2\right\}\exp\left\{\frac{t^2}{2n\sigma^2}\right\}\delta\left(\sum_{k=0}^{n-1}y_k - t\right). \end{aligned}$$

We see that, conditioned on $Z = T(Y) = t$, the pdf of the observations does not depend on the unknown parameter θ. Observing $Z = T(Y) = t$ squeezes all of the information about the unknown parameter out of the observations. Hence, $T(y) = \sum_{k=0}^{n-1} y_k$ is a sufficient statistic.

Sufficiency and Minimal Sufficiency

Definition

$T : \mathcal{Y} \mapsto \Delta$ is a **sufficient statistic** for the family of parameterized pdfs $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$ if the distribution of the random observation conditioned on $T(Y)$, i.e. $p_Y\left(y \,|\, T(Y) = t\,;\,\theta\right)$, does not depend on θ for all $\theta \in \Lambda$ and all $t \in \Delta$.

Intuitively, a sufficient statistic summarizes the information contained in the observation about the unknown parameter. Knowing $T(y)$ is as good as knowing the full observation y when we wish to estimate θ.

Definition

$T : \mathcal{Y} \mapsto \Delta$ is said to be **minimal sufficient** for the family of pdfs $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$ if it is a function of every other sufficient statistic for this family of pdfs.

Intuitively, a minimal sufficient statistic is the most concise summary of the observation. These can be hard to find and don't always exist.

Example: Sufficiency

Suppose $\theta \in \mathbb{R}$ and we get a vector observation $y \in \mathbb{R}^n$ distributed as

$$p_Y(y\,;\,\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{k=0}^{n-1} (y_k - \theta)^2 \right\} = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{-\|y - 1\theta\|_2^2}{2\sigma^2} \right\}.$$

Is $T(y) = y = [y_0, \dots, y_{n-1}]^\top$ sufficient? Intuitively, it should be. To gain some experience with the definition, however, let's check. What happens to $p_Y(y\,;\,\theta)$ when we condition on an observation $Y = [y_0, \dots, y_{n-1}]^\top$?

Neyman-Fisher Factorization Theorem

Guessing and checking sufficient statistics isn't very satisfying. We would prefer a **procedure** for finding sufficient statistics.

Theorem (Fisher 1920, Neyman 1935)

A statistic T is sufficient for θ if and only if there exist functions g_θ and h such that the parameterized pdf of the observation can be factored as

$$p_Y(y\,;\,\theta) \quad = \quad g_\theta(T(y))h(y)$$

for all $y \in \mathcal{Y}$ and all $\theta \in \Lambda$.

The proof of this theorem is in Appendix 5A of Kay I.

Note that $h(y)$ can't be a function of θ and $g_\theta(T(y))$ must only be a function of θ and $T(y)$.

Example: Neyman-Fisher Factorization Theorem

Suppose $\theta \in \mathbb{R}$ and

$$p_Y(y\,;\,\theta) \;=\; \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta)^2\right\}.$$

Let $T(y) = \frac{1}{n}\sum_{k=0}^{n-1} y_k = \bar{y}$. We already know this is a sufficient statistic but let's try the factorization.

$$\begin{aligned} p_Y(y\,;\,\theta) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{\frac{-n}{2\sigma^2}\left(\frac{1}{n}\sum_{k=0}^{n-1} y_k^2 - 2\theta y_k + \theta^2\right)\right\} \\ &= \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{\frac{-n}{2\sigma^2}\left(\theta^2 - 2\theta\bar{y}\right)\right\}}_{g_\theta(T(y))} \underbrace{\exp\left\{\frac{-1}{2\sigma^2}\sum_{k=0}^{n-1} y_k^2\right\}}_{h(y)} \end{aligned}$$

Flashback: Sufficient Statistic For Simple Binary HT

Suppose we have a simple binary hypothesis testing problem:

$$Y \sim p_Y(y\,;\,\theta) = \begin{cases} p_0(y) & \theta = 0 \\ p_1(y) & \theta = 1 \end{cases}$$

Let

$$\begin{aligned} T(y) &= \frac{p_1(y)}{p_0(y)} \\ g_\theta(T(y)) &= \theta T(y) + (1-\theta) \\ h(y) &= p_0(y) \end{aligned}$$

Then it is easy to show that $p_Y(y\,;\,\theta) = g_\theta(T(y))h(y)$. Hence the likelihood ratio $L(y) = \frac{p_1(y)}{p_0(y)}$ is a sufficient statistic for simple binary hypothesis testing problems.

Completeness of a Family of PDFs

Definition

The family of pdfs $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$ is said to be complete if the condition $\mathrm{E}_\theta\left[f(Y)\right] = 0$ for all θ in Λ implies that $\mathrm{Prob}_\theta[f(Y) = 0] = 1$ for all θ in Λ. Note that $f : \mathcal{Y} \mapsto \mathbb{R}$ can be any function.

To get some intuition, consider the case where $\mathcal{Y} = \{y_0, \dots, y_{L-1}\}$ is a finite set. Then

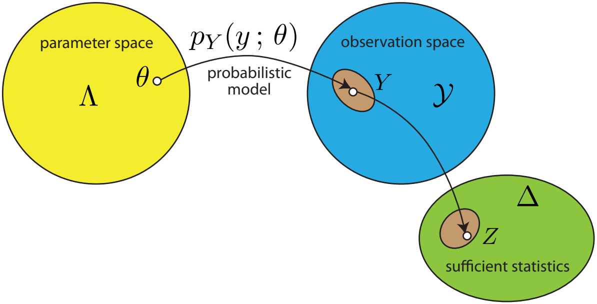
$$\begin{aligned}\mathrm{E}_\theta\left[f(Y)\right] &= \sum_{\ell=0}^{L-1} f(y_\ell)\mathrm{Prob}_\theta(Y = y_\ell)\\ &= f^\top(y) P_\theta\end{aligned}$$

For a fixed θ it is certainly possible to find a non-zero f such that $\mathrm{E}_\theta\left[f(Y)\right] = 0$. But we have to satisfy this condition for all $\theta \in \Lambda$, i.e. we need a vector $f(y)$ that is **orthogonal to the all members of the family** of vectors $\{P_\theta\,;\,\theta \in \Lambda\}$. If the only such vector that satisfies the condition $\mathrm{E}_\theta\left[f(Y)\right] = 0$ for all $\theta \in \Lambda$ is $f(y_0) = \cdots = f(y_{L-1}) = 0$, then the family $\{P_\theta\,;\,\theta \in \Lambda\}$ is complete.

Complete Sufficient Statistics

Definition

Suppose that T is a sufficient statistic for the family of pdfs $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$. Let $p_Z(z\,;\,\theta)$ denote the distribution of $Z = T(Y)$ when the parameter is θ. If the family of pdfs $\{p_Z(z\,;\,\theta)\,;\,\theta \in \Lambda\}$ is complete, then T is said to be a complete sufficient statistic for the family $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$.



Example: Complete Sufficient Statistic

Suppose $\theta \in \mathbb{R}$ and

$$p_Y(y\,;\,\theta) \;=\; \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta)^2\right\}.$$

Let $T(y) = \frac{1}{n}\sum_{k=0}^{n-1} y_k$. We know this is a sufficient statistic but is it complete?

We know the distribution of $T(Y) = \bar{Y}$ is $\mathcal{N}(\theta, \sigma^2/n)$. We require $\mathrm{E}_\theta[f(\bar{Y})] = 0$ for all $\theta \in \Lambda$, i.e.

$$s(\theta) = \int_{-\infty}^{\infty} f(x)\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-n(x-\theta)^2}{2\sigma^2}\right\}dx = 0 \text{ for all } \theta \in \Lambda$$

Is there any non-zero $f : \mathbb{R} \mapsto \mathbb{R}$ that can do this? Suppose $\theta = 0$. Lots of functions will do this, e.g. $f(x) = x$, $f(x) = \sin(x)$, etc. But we need $s(\theta) = 0$ for all $\theta \in \Lambda$.

Example: Complete Sufficient Statistic (continued)

$$s(\theta) = \int_{-\infty}^{\infty} f(x) \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-n(x-\theta)^2}{2\sigma^2} \right\} dx = 0 \text{ for all } \theta \in \Lambda$$

$$\Leftrightarrow \quad s(\theta) = \int_{-\infty}^{\infty} f(x) \exp\left\{ (\theta - x)^2 \right\} dx = 0 \text{ for all } \theta \in \Lambda$$

But this is just the convolution of $f(x)$ with a Gaussian pulse.

Recall that convolution in the "time domain" is multiplication in the "frequency domain". Hence, if $S(\omega)$ is the Fourier transform of $s(\theta)$, then

$$\Leftrightarrow \quad S(\omega) = F(\omega)G(\omega) = 0 \text{ for all } \omega$$

where $G(\omega)$ is the Fourier transform of the Gaussian pulse. Note that $G(\omega)$ is itself Gaussian and therefore positive for all ω. Hence, the only way to force $S(\omega) \equiv 0$ is to have $F(\omega) \equiv 0$. Hence the only solution to $\mathrm{E}_\theta[f(\bar{Y})] = 0$ for all $\theta \in \Lambda$ is $f(x) \equiv 0$ for all x and, consequently, $T(y)$ is a complete sufficient statistic.

Example: Incomplete Sufficient Statistic

Suppose $\theta \in \mathbb{R}$ and you would like to estimate θ from a scalar observation $Y = \theta + W$ where $W \sim \mathcal{U}\left[-\frac{1}{2}, \frac{1}{2}\right]$.

An obvious sufficient statistic then is $T(y) = y$. But is it complete?

Since $T(Y) = Y$, we require $\mathrm{E}_\theta[f(Y)] = 0$ for all $\theta \in \Lambda$, i.e.

$$s(\theta) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} f(x)\,dx = 0 \text{ for all } \theta \in \Lambda$$

Is there any non-zero $f : \mathbb{R} \mapsto \mathbb{R}$ that can do this?

How about $f(x) = \sin(2\pi x)$? This definitely forces $s(\theta) = 0$ for all $\theta \in \Lambda$. Just need to confirm $\mathrm{Prob}[f(Y) = 0] < 1$ for at least one $\theta \in \Lambda$.

Since we found a non-zero $f(x)$ that forced $\mathrm{E}_\theta[f(Y)] = 0$ for all $\theta \in \Lambda$, we can say that $T(y) = y$ is not complete.

Completeness Theorem for Exponential Families

Theorem

Suppose $\mathcal{Y} = \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$, *and*

$$p_Y(y\,;\,\theta) \;=\; a(\theta)\exp\left\{\sum_{\ell=1}^{m}\theta_\ell T_\ell(y)\right\}h(y)$$

where $a, T_1, \dots, T_m$, *and* h *are all real-valued functions. Then* $T(y) = [T_1(y), \dots, T_m(y)]^\top$ *is a complete sufficient statistic for the family* $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$ *if* Λ *contains an* m*-dimensional rectangle.*

Remarks:

- The technical detail about the m-dimensional rectangle ensures that Λ is not missing any dimensions in $\mathbb{R}^m$, e.g. Λ is not a two-dimensional plane in $\mathbb{R}^3$.
- Main idea of proof is similar to how we showed completeness in the Gaussian example. See Poor pp. 165-166 and/or Lehmann 1986.

Rao-Blackwell-Lehmann-Sheffe Theorem

Theorem

If $\hat{g}(y)$ *is any unbiased estimator of* θ *and* T *is a sufficient statistic for the family* $\{p_Y(y\,;\,\theta)\,;\,\theta \in \Lambda\}$, *then*

$$\tilde{g}[T(y)] \quad := \quad \mathrm{E}_\theta\left[\hat{g}(Y)\,|\,T(Y) = T(y)\right]$$

is

- *A valid estimator of* θ *(not a function of* θ*)*
- *An unbiased estimator of* θ.
- *Of lesser or equal variance than that of* $\hat{g}(y)$ *for all* $\theta \in \Lambda$

Additionally, if T *is complete, then* $\tilde{g}[T(y)]$ *is an MVU estimator of of* θ.

Example: Estimating a Constant in White Gaussian Noise

Suppose $\theta \in \mathbb{R}$ and

$$p_Y(y\,;\,\theta) \;=\; \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{k=0}^{n-1} (y_k - \theta)^2 \right\}.$$

Let $T(y) = \frac{1}{n}\sum_{k=0}^{n-1} y_k$. We know this is a complete sufficient statistic. Let's apply the RBLS theorem to find the MVU estimator...

- We could choose the unbiased estimator $\hat{g}(y) = y_0$.
- Now we need to compute

$$\tilde{g}[T(y)] := \mathrm{E}_\theta\left[\hat{g}(Y)\,|\,T(Y) = T(y)\right] = \mathrm{E}_\theta\left[Y_0\,\middle|\,\frac{1}{n}\sum Y_k = \frac{1}{n}\sum y_k\right]$$

- To solve this, we can use a standard formula for the conditional expectation of a jointly Gaussian random variable...

Example (continued)

- Suppose $Z = [X, Y]^\top$ is jointly Gaussian distributed. We know that

$$\mathrm{E}[X|Y=y] \;=\; \mathrm{E}[X] + \frac{\mathsf{cov}(X,Y)}{\mathsf{var}(Y)}(y - \mathrm{E}[Y]).$$

- In our problem, letting $\bar{Y} = \frac{1}{n}\sum_{k=0}^{n-1} Y_k$, we can use this result to write

$$\begin{aligned}\mathrm{E}_\theta\left[Y_0 \,|\, \bar{Y} = t\right] &= \mathrm{E}_\theta[Y_0] + \frac{\mathsf{cov}_\theta(Y_0, \bar{Y})}{\mathsf{var}_\theta(\bar{Y})}\left(\frac{1}{n}\sum y_k - \mathrm{E}_\theta[\bar{Y}]\right) \\ &= \theta + \frac{\sigma^2}{\sigma^2}\left(\frac{1}{n}\sum y_k - \theta\right) \\ &= \frac{1}{n}\sum y_k\end{aligned}$$

- Hence $\hat{\theta}_{\mathsf{mvu}}(y) = \frac{1}{n}\sum y_k$ is an MVU estimator of θ.

Conclusions

- To approach non-random parameter estimation problems, we had to restrict our attention to the class of unbiased estimators.
- Under the squared error cost assignment, the performance of these unbiased estimators is measured by the variance of the estimates.
- Rao-Blackwell-Lehmann-Sheffe theorem establishes a procedure for finding minimum variance unbiased (MVU) estimators.
- Subtle concepts will require some practice:
  - Sufficiency (Neyman-Fisher factorization theorem)
  - Completeness
- Following RBLS doesn't guarantee you will find an MVU estimator:
  - It can be difficult/impossible to find a complete sufficient statistic.
  - It is often difficult to check completeness.
  - Computing the conditional expectation can be intractable.
- Other techniques may be useful for checking if an estimator is MVU.
- Further restrictions on the class of estimators also facilitate analysis.