Chapter 3.(b) Minimum-variance parameter estimation

Non-random parameter estimation: MVUE

- No prior on $\Theta$.
- How to evaluate performance?
  - Cant average conditional risk without the prior
  - $\hat{\theta}(y) = \theta_0$ may work well only around $\theta_0$.
  - Seek estimators suitable for all $\theta$.
  - Restrict the search to unbiased estimator:
    $$E[\hat{\theta}(y)] = \theta, \quad \forall \theta \in \Lambda$$
- Minimize the MSE risk:
  $$R_{\theta}(\hat{\theta}) = E_Y[(\hat{\theta} - \theta)^2]$$
- For unbiased case the above equals to the variance.
- Minimum variance unbiased estimator (MVUE).
- If linear, then it is called best linear unbiased estimate (BLUE).
Sufficient statistics

- Can we summarize the observations without penalty in estimating $\theta$?
- Observation: $Y$, Statistic: $T(Y)$.
- **Sufficient statistic:** $T(Y)$ is sufficient statistic for estimating $\theta$, if
  
  $$P_\theta(Y = y | T(Y) = t)$$
  
  is not a function of $\theta$.

**Minimal sufficient statistic:** A sufficient statistic $T$ is minimal, if it is a function of every other sufficient statistic. (May be hard to find.)

Example

- Consider $Y_k = \Theta + N_k$, $N_k \sim \mathcal{N}(0,1)$ i.i.d. with $T(Y) = \frac{1}{n} \sum_{k=1}^{n} Y_k$. 
The Factorization Theorem

**Theorem**

*Given a family of densities \( \{p_\theta, \theta \in \Lambda\} \), a statistic \( T \) is sufficient if and only if there exists functions \( g_\theta(\cdot) \) and \( h(\cdot) \) such that, for all \( y \in \Gamma \) and \( \theta \in \Lambda \), \( p_\theta(y) = g_\theta[T(y)]h(y) \).*

**Proof:**

Example: Likelihood ratio

- Consider hypothesis testing \( \Lambda = \{0, 1\} \) with densities \( p_0, p_1 \).
The Rao-Blackwell Theorem

- Consider estimating $g(\theta)$, a function of $\theta$.
- **Idea:** Use sufficient statistic to improve an unbiased estimator.

**Theorem**

Suppose $\hat{g}(y)$ is an unbiased estimate of $g(\theta)$ and $T$ is sufficient for $\theta$. Define

$$\tilde{g}(T(y)) = E_{\theta}[\hat{g}(Y) | T(Y) = T(y)].$$

Then,

a) $\tilde{g}(T(Y))$ is also an unbiased estimate of $g(\theta)$, and

b) $\text{var}_{\theta}(\tilde{g}(T(Y))) \leq \text{var}_{\theta}(\hat{g}(Y))$, with equality if and only if $P_{\theta}(\hat{g}(Y) = \tilde{g}(T(Y))) = 1$.

**Proof:**

- a) Unbiasedness:

- b) Variance:
Corollary to Rao-Blackwell Theorem

Corollary

If $T$ is a sufficient statistic for $\theta$ and there exists a unique function of $T$ that is an unbiased estimator of $g(\theta)$, then that function of $T$ is an MVUE of $g(\theta)$.

Proof:

Completeness

- A family of distributions $\{P_\theta, \theta \in \Lambda\}$ is **complete** if $E_\theta[f(Y)] = 0, \forall \theta \in \Lambda$ implies that $f(Y) = 0$ with probability 1, $\forall \theta \in \Lambda$.
- Analogy: Set of vectors in $\mathbb{R}^n$ is called complete if they span $\mathbb{R}^n$.
- Discrete example:
Examples on Completeness

\[
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}
\theta \\
\theta
\end{bmatrix}, \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}\right),
\]
\[f(y) = y_1 - y_2.\]

\[
Y \sim \mathcal{N}(\theta, 1) \text{ and } \theta \in \mathcal{N}(0, 1).
\]

Completeness and Sufficiency

Let \( T \) be sufficient statistic for \( \theta \), and \( T(Y) \sim Q_\theta \) when \( Y \sim P_\theta \). If \( \{Q_\theta, \theta \in \Lambda\} \) is complete, then \( T \) is said to be a complete sufficient statistic.

Theorem

Any unbiased estimator that is a function of a complete sufficient statistic is MVUE.

Proof:
A recipe for MVUE

- Finding MVUE:
  - Find a complete sufficient statistic for $\theta$.
  - Find an unbiased estimator $\hat{g}(y)$ of $g(\theta)$.
  - Then, $\tilde{g}(T(y)) = E_{\theta} [\hat{g}(Y) | T(Y) = T(y)]$ is an MVUE of $g(\theta)$.
- First step is easy for exponential families.

Exponential Families

- $\{P_\theta, \theta \in \Lambda\}$ is said to be an exponential family if there are real-valued functions $C, Q_1, \ldots, Q_m, T_1, \ldots, T_m, h$ such that
  $$p_\theta(y) = C(\theta) \exp \left\{ \sum_{l=1}^m Q_l(\theta) T_l(y) \right\} h(y)$$
  for all $\theta \in \Lambda$ and $y \in \Gamma$.
- Completeness theorem for exponential families: Suppose $\Gamma = \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$, and, for each $\theta$,
  $$p_\theta(y) = C(\theta) \exp \left\{ \sum_{l=1}^m \theta_l T_l(y) \right\} h(y)$$
  with real-valued functions $C, T_1, \ldots, T_m, h$. Then, $T(y) = [T_1(y), \ldots, T_m(y)]$ is a complete sufficient statistic for $\{P_\theta, \theta \in \Lambda\}$ if $\Lambda$ contains a $m$-dimensional rectangle.
- Note: Complete sufficient statistics are minimal.
Example

- Exponential families:
  Gaussian, Laplacian, Poisson, binomial, geometric distributions.

- Complete sufficient statistic:
  Consider \[ Y_1, Y_2 \sim \mathcal{N}\left(\begin{bmatrix} \theta \\ \theta \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \theta \in \mathbb{R}. \]

MVUE Example 1

- \[ Y_k = N_k + \mu s_k \text{ for } k = 1, \ldots, n \]
  - \[ N \sim \mathcal{N}(0, \sigma^2 I). \]
  - \( s_k \) are known \((s_1 \neq 0)\)
  - \( \sigma^2 \) known \((\sigma^2 > 0)\)
  - Objective: Estimate \( \mu \).

- Consider
  \[ p_{\mu}(y) = \]

- We have
  - \( \theta_1 = \)
  - \( T(y) = \)
  - \( h(y) = \)
  - \( C(\overline{y}) = \)

- \( T \) is a complete sufficient statistic.
MVUE Example 1

- Consider $\hat{g}(y) = \frac{Y_1}{s_1}$
- Define $\tilde{g}(T(y)) = E_{\theta_1}[\frac{Y_1}{s_1} | T(Y) = T(y)]$

MVUE Example 1

- Consider $A$ and $B$ are jointly Gaussian.

\[
\begin{bmatrix}
A \\
B
\end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix}
\mu_A \\
\mu_B
\end{bmatrix}, \begin{bmatrix}
\Sigma_A & \Sigma_{AB} \\
\Sigma_{AB}^T & \Sigma_B
\end{bmatrix} \right)
\]

- Then, $B|A \sim \mathcal{N}(\mu_B + \Sigma_{AB} \Sigma_A^{-1}(A - \mu_A), \Sigma_B - \Sigma_{AB} \Sigma_A^{-1} \Sigma_{AB})$
MVUE Example 1

Recall the Bayes estimator

\[ \hat{\mu}_B = \gamma \left( \frac{s^T y}{||s||^2} \right) + (1 - \gamma) \mu \]

MVUE Example 2

Objective: Estimate \( \mu, \sigma^2 \).

Consider

\[ p_{\mu,\sigma^2}(y) = \]

We have

- \( \theta_1 = \)
- \( \theta_2 = \)
- \( T_1(y) = \)
- \( T_2(y) = \)
- \( h(y) = \)
- \( C(\theta_1, \theta_2) = \)

Is \( T = [T_1, T_2] \) a complete sufficient statistic for \( \theta \)?
MVUE Example 2

Need to estimate $\mu = g_1(\theta) = \frac{-\theta_1}{2\theta_2}$, and $\sigma^2 = g_2(\theta) = \frac{-1}{2\theta_2}$. 