E2 : Minimum Variance Unbiased Estimation

Goal: Find good estimators of unknown deterministic parameters for a generative statistical model of the data from random observations.

Ex] $x[n] = A + w[n]$ where A is an unknown constant and $w[n]$ is a zero-mean noise process. Estimate A using observations $x[n]$.

Unbiased Estimators

Defn: An estimator $\hat{\theta}$ for a parameter θ is unbiased iff $E[\hat{\theta}] = \theta \quad \forall \theta$.

Ex] Let $x[n] = A + w[n]$, $n = 0, 1, \dots, N-1$ where $A \in \mathbb{R}$ is the parameter to be estimated and $w[n]$ is WGN (here we have a DC-level with AWGN).

Consider $\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$, the sample average as an estimator. We have

$$E[\hat{A}] = E\left[\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right] = \frac{1}{N} \sum_{n=0}^{N-1} E[x[n]] = A$$

$\therefore$ The sample average is an unbiased estimator here.

Note: An unbiased estimator is not necessarily a good one.

Fusion:

If several unbiased estimators $\{\hat{\theta}_1, \dots, \hat{\theta}_M\}$ are available, then a weighted average $\hat{\theta} = \sum_{m=1}^{M} \alpha_m \hat{\theta}_m$ with some good selection of α according to an optimality criterion could yield unbiased $\hat{\theta}$, since $E[\hat{\theta}] = \sum_{m=1}^{M} \alpha_m E[\hat{\theta}_m] = \theta$, with ~~lower~~ better performance than each individual estimator.

For instance if $\theta \in \mathbb{R}$ (scalar) and $\mathrm{Cov}\left[\begin{pmatrix}\hat{\theta}_1\\ \vdots \\ \hat{\theta}_M\end{pmatrix}\right] = C$, then $\mathrm{Var}(\hat{\theta}) = \alpha^T C \alpha$. Choosing α to be the ~~eigenvector of C with the smallest eigenvalue~~ minimizer of $\alpha^T C \alpha$ subject to $\alpha^T \mathbb{1} = 1$ would yield an unbiased $\hat{\theta}$ with the smallest variance in this linear fusion rule.

$$\min_{\alpha} \frac{1}{2} \alpha^T C \alpha \quad \text{s.t.} \quad \alpha^T \mathbb{1} = 1$$

$$\mathcal{L}(\alpha) = \frac{1}{2} \alpha^T C \alpha - \lambda (\alpha^T \mathbb{1} - 1)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha^T} = \frac{1}{2} 2 C \alpha - \lambda \mathbb{1} = 0 \quad \text{Multiplying from left w/ } \alpha^T$$

substitute $\Rightarrow \frac{1}{\cancel{2}} \cancel{2}\, \alpha^T C \alpha - \lambda \alpha^T \mathbb{1} = \alpha^T 0$

$$C\alpha = (\alpha^T C \alpha) \mathbb{1} \qquad \Leftrightarrow \boxed{\alpha^T C \alpha = \lambda}$$

The solution for α that satisfies this equation and has the smallest λ value is the optimal linear fusion solution

A suboptimal unbiased choice (if we take $var(\hat{\theta})$ as the optimality objective) would be $\hat{\theta} = \frac{1}{M}\sum_{m=1}^{M} \hat{\theta}_m$.

This is $\alpha = \frac{\mathbf{1}}{M}$. Then $Var(\hat{\theta}) = \alpha^T C \alpha = \frac{\mathbf{1}^T C \mathbf{1}}{M^2}$.

If $C = \sigma^2 I$ (i.e. all estimators are uncorrelated and have identical variance), then $Var(\hat{\theta}) = \frac{\sigma^2}{M}$.

This shows that fusing uncorrelated estimators have the potential of significantly reducing estimation variance.

Minimum Variance Criterion

Consider estimator $\hat{\theta} \in \mathbb{R}$ for parameter $\theta \in \mathbb{R}$. The mean squared error (MSE) of this estimator is

$$
\begin{aligned}
MSE(\hat{\theta}) &\triangleq E[(\hat{\theta}-\theta)^2] \\
&= E\{[(\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta)]^2\} \\
&= Var(\hat{\theta}) + [E[\hat{\theta}] - \theta]^2 \\
&= Var(\hat{\theta}) + bias^2(\hat{\theta})
\end{aligned}
$$

Ex) $x[n] = A + w[n]$ $n = 0, \ldots, N-1$ and $w[n]$ is WGN.

Consider $\hat{A} = a \frac{1}{N} \sum_{n=0}^{N-1} x[n]$ for some constant a.

$$MSE(\hat{A}) = var(\hat{A}) + bias^2(\hat{A}) = \frac{a^2\sigma^2}{N} + (a-1)^2 A^2$$

$$\frac{\partial MSE(\hat{A})}{\partial a} = \frac{2a\sigma^2}{N} + 2(a-1)A^2 = 0 \Rightarrow a_* = \frac{A^2}{A^2 + (\sigma^2/N)}$$

Depends on unknown A

NOT REALIZABLE

The minimum MSE estimator is in general not realizable. For practical problems, one approach is to constrain the bias to be zero and minimize the variance. This has the nice side effect of concentrating the pdf of $\hat{\theta}-\theta$ around zero, thus making large errors unlikely for many cases.

Existence of the Minimum Variance Unbiased Estimator

Fact: In general, the MVU estimator does NOT exist.

Ex) $x[0] \sim \underbrace{\mathcal{N}}_{\text{normal pdf}}(\underbrace{\theta}_{\text{mean}}, \underbrace{1}_{\text{var.}})$

$$x[1] \sim \begin{cases} \mathcal{N}(\theta, 1) & \text{if } \theta \geq 0 \\ \mathcal{N}(\theta, 2) & \text{if } \theta < 0 \end{cases}$$

Consider two unbiased estimators: $\hat{\theta}_1 = \frac{1}{2}(x[0] + x[1])$

$$\hat{\theta}_2 = \frac{2}{3}x[0] + \frac{1}{3}x[1]$$

We have $\text{var}(\hat{\theta}_1) = \begin{cases} 18/36 & \text{if } \theta \geq 0 \\ 27/36 & \text{if } \theta < 0 \end{cases}$ (after some algebra)

$$\text{var}(\hat{\theta}_2) = \begin{cases} 20/36 & \text{if } \theta \geq 0 \\ 24/36 & \text{if } \theta < 0 \end{cases}$$

$\hat{\theta}_1$ is better for $\theta \geq 0$ and $\hat{\theta}_2$ is better for $\theta < 0$.

For $\theta \geq 0$, $\hat{\theta}_1$ is ~~the~~ MVU (but not $\forall \theta \in \mathbb{R}$). For $\theta < 0$ $\hat{\theta}_2$ is MVU (but not $\forall \theta \in \mathbb{R}$). No estimator exists which ~~uniformly~~ uniformly ($\forall \theta \in \mathbb{R}$) has minimum variance.

Finding the MVU Estimator

Even if a MVU estimator exists, we may not be able to find it easily. Some possible procedures are:

1) Determine the Cramer-Rao lower Bound (CRLB) and check if some estimator satisfies it. (Chapters 3, 4)
2) Apply the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem. (Chapter 5)
3) Further restrict the class of estimators to be not only unbiased but also linear. (Chapter 6).

Extension to a Vector Parameter

Let $\theta \in \mathbb{R}^p$. Then $\hat{\theta} \in \mathbb{R}^p$ is unbiased iff $E[\hat{\theta}] = \theta \;\; \forall \theta$.

$\hat{\theta}$ is MVU if each $\hat{\theta}_i$ is MVU for $i = 1, \ldots, p$.

Suggested Problems: 1, 2, 3, 6, 9, 10