ECE531 Lecture 10a: Best Linear Unbiased Estimation

D. Richard Brown III

Worcester Polytechnic Institute

06-April-2011

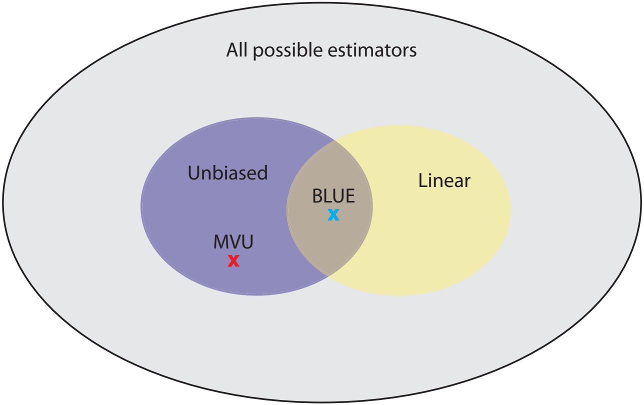
Introduction

- In this lecture, we continue our study of **unbiased** estimators of **non-random parameters** under the **squared error** cost function.
- Squared error: Estimator variance determines performance.
- We seek to find the minimum variance unbiased (MVU) estimator.
- So far, we have two approaches to finding MVU estimators:
  1. Rao-Blackwell-Lehmann-Sheffe
  2. Guess and check with respect to the Cramer-Rao lower bound
- Both approaches can be difficult, as you've seen.
- A common approach often used in practical implementations: further restrict our attention to unbiased linear estimators, i.e.

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

  where $A \in \mathbb{R}^{n \times m}$ is a linear mapping from observations to estimates.
- We now seek to find the "best linear unbiased estimator" (BLUE).

Best Linear Unbiased Estimator



- In general, the BLUE will not be the same as the MVU estimator.
- What can we say about the squared error performance of the BLUE with respect to the MVU?
- When will BLUE=MVU?

Example 1

Suppose we have random observations given by

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

where $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ with $\theta \in \mathbb{R}$. What is the MVU estimator for θ?

What is the BLUE estimator for θ?

Example 2

Suppose we have random observations given by

$$Y_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(0, \beta) \qquad k = 0, \dots, n-1$$

and we wish to estimate the mean $\theta = \beta/2$. What is the MVU estimator for θ?

We can confirm that $T(y) = \max y$ is a complete sufficient statistic for this problem (See Kay I: Example 5.8). Grinding through the RBLS yields

$$\hat{\theta}_{\mathsf{MVU}}(y) = \frac{N+1}{2N} T(y) = \frac{N+1}{2N} \max y$$

Does MVU=BLUE in this case?

How can we find the BLUE?

Finding the BLUE: Problem Setup

Denote the BLUE estimator as $\hat{\theta}_{\mathsf{BLUE}}(y) = \bar{A}y$ where $\bar{A} \in \mathbb{R}^{n \times m}$. We wish to solve

$$\bar{A} = \arg \min_{A \in \mathbb{R}^{n \times m}} \mathrm{trace}\left[\mathrm{cov}\left\{AY\right\}\right] \tag{1}$$

subject to the constraint that $\mathrm{E}\left\{\bar{A}Y\right\} = \theta$ for all $\theta \in \Lambda$.

Recall that the trace of a matrix is the sum of its diagonal elements. Hence, we seek to find the linear unbiased estimator that minimizes the sum of the variances.

Finding the BLUE: The Constraint (part 1)

Let's look at the unbiased constraint first. Since $\bar{A}$ is a constant and linear, the unbiased constraint can be written as

$$\bar{A}\mathrm{E}\{Y\} = \theta.$$

- Example 1: Suppose you have scalar θ and get observations $Y_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$ for $k = 0, \dots, n-1$. What does the unbiased constraint imply about $\bar{A}$?
- Example 2: Suppose you have scalar θ and get observations $Y_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-\theta, \theta)$ for $k = 0, \dots, n-1$. What does the unbiased constraint imply about $\bar{A}$?

Bottom line: Lots of problems make sense in the BLUE context, but not every problem. You should confirm that it is possible to have an unbiased linear estimator before proceeeding.

Finding the BLUE: The Constraint (part 2)

The unbiased constraint $\bar{A}\mathrm{E}\{Y\} = \theta$ can be satisfied if and only if

$$\mathrm{E}\{Y\} = H\theta.$$

for some known $H \in \mathbb{R}^{n\times m}$ with full column rank, i.e. H must have m linearly independent columns. In other words, $\mathrm{E}\{Y\}$ must be linear in θ for some known H with full column rank ($H \neq 0$ for scalar parameters).

The proof of this result follows from the fact that there exists a "left inverse" $A \in \mathbb{R}^{m\times n}$ of H such that $AH = I$ if and only if H has full column rank.

- If the left inverse does exist, then the unbiased constraint can be satisfied since there is at least one $A \in \mathbb{R}^{m\times n}$ such that $A\mathrm{E}\{Y\} = AH\theta = \theta$.
- If the left inverse does not exist, then the unbiased constraint can't be satisfied since $A\mathrm{E}\{Y\} \neq \theta$ for all $A \in \mathbb{R}^{m\times n}$.

Examples

Suppose you get observations $Y_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(\theta_1, \theta_2)$ for $k = 0, \dots, n-1$. Can we find an H with full column rank such that

$$\mathrm{E}\{Y\} = H\theta?$$

Suppose you get observations $Y_k \sim \mathcal{U}(\theta_1, k\theta_2)$ for $k = 0, \dots, n-1$. Can we find an H with full column rank such that

$$\mathrm{E}\{Y\} = H\theta?$$

Finding the BLUE: The Minimization (part 1)

Recall that we wish to solve

$$\bar{A} = \arg \min_{A \in \mathbb{R}^{n \times m}} \operatorname{trace}\left[\operatorname{cov}\{AY\}\right] \tag{2}$$

subject to the unbiased constraint $AH = I$. We can compute

$$\begin{aligned}
\operatorname{cov}\{AY\} &= \mathrm{E}\left\{[AY - \mathrm{E}(AY)]\,[AY - \mathrm{E}(AY)]^\top\right\} \\
&= A\mathrm{E}\left\{[Y - \mathrm{E}(Y)]\,[Y - \mathrm{E}(Y)]^\top\right\} A^\top \\
&= A\operatorname{cov}\{Y\}\, A^\top \\
&= ACA^\top
\end{aligned}$$

where $C := \operatorname{cov}\{Y\}$ is the covariance of the observations (assumed to be known), possibly parameterized by θ.

Finding the BLUE: The Minimization (part 2)

Now we wish to solve

$$\bar{A} = \arg \min_{A \in \mathbb{R}^{n \times m}} \text{trace}\left(ACA^\top\right). \tag{3}$$

subject to the unbiased constraint $AH = I$. An aside: What would A be if we didn't have the constraint?

Recall that the trace of a matrix is the sum of the diagonal elements. Hence, denoting e_i as the i^{th} standard basis vector, we can write

$$\text{trace}\left(ACA^\top\right) = \sum_i e_i^\top ACA^\top e_i = \sum_i a_i^\top C a_i$$

where $a_i^\top$ is the i^{th} row of the A matrix, i.e.

$$A = \begin{bmatrix} a_0^\top \\ \vdots \\ a_{m-1}^\top \end{bmatrix}.$$

Finding the BLUE: The Minimization (part 3)

Now we wish to solve

$$\bar{A} = \arg \min_{A \in \mathbb{R}^{n \times m}} \sum_i a_i^\top C a_i. \tag{4}$$

subject to the unbiased constraint $AH = I$. Note that each element in this sum can be **minimized separately** since the first element only depends on a_0, the second element only depends on a_1, and so on. These minimization problems are linked by their constraints, however.

So, for each $i = 0, 1, \ldots, m-1$, we can instead solve

$$\bar{a}_i = \arg \min_{a_i \in \mathbb{R}^n} a_i^\top C a_i. \tag{5}$$

subject to $AH = I$. How do we solve these sort of problems?

Finding the BLUE: The Minimization (part 4)

We can solve the i^{th} subproblem

$$\bar{a}_i = \arg \min_{a_i \in \mathbb{R}^n} a_i^\top C a_i. \tag{6}$$

subject to $AH = I$ by using the Lagrange multiplier method with multiple constraints.

Let $f(a_i) = a_i^\top C a_i$ and let $g_j(a_i) = a_i^\top h_j - \delta_{ij}$ where h_j is the j^{th} column of H and δ_{ij} is the Kronecker delta function. We wish to minimize $f(a_i)$ subject to the constraints $g_j(a_i) = 0$ for all j. To do this, we solve the system of equations

$$\nabla_{a_i} f(a_i) = \sum_j \lambda_j \nabla_{a_i} g_j(a_i)$$

$$g_j(a_i) = 0 \qquad \forall j$$

Finding the BLUE: The Minimization (part 5)

Substituting in for $f(a_i)$ and $g_j(a_i)$, we have

$$\nabla_{a_i}(a_i^\top C a_i) = \sum_j \lambda_j \nabla_{a_i}(a_i^\top h_j - \delta_{ij})$$
$$a_i^\top h_j - \delta_{ij} = 0 \qquad \forall j$$

and doing the gradients yields

$$2Ca_i = \sum_j \lambda_j h_j$$
$$a_i^\top h_j - \delta_{ij} = 0 \qquad \forall j.$$

This can be put into a more compact matrix-vector notation as

$$2Ca_i = H\lambda$$
$$a_i^\top H = e_i^\top \qquad \forall j.$$

where $\lambda \in \mathbb{R}^m$ and e_i is the i^{th} standard basis vector.

Finding the BLUE: The Minimization (part 6)

We have

$$2Ca_i = H\lambda$$
$$a_i^\top H = e_i^\top.$$

The first equation implies

$$a_i = \frac{1}{2}C^{-1}H\lambda. \tag{7}$$

We just need to solve for $\lambda \in \mathbb{R}^m$ by using the constraint.

The constraint equation can be equivalently written as $H^\top a_i = e_i$. Hence, we can multiply (7) by $H^\top$ to write

$$H^\top a_i = \frac{1}{2}H^\top C^{-1}H\lambda = e_i.$$

The quantity $H^\top C^{-1}H$ has full rank, hence we can write

$$\lambda = 2(H^\top C^{-1}H)^{-1}e_i$$

Finding the BLUE: The Minimization (part 7)

We plug this result back into (7) to get the solution to the i^{th} subproblem as

$$\bar{a}_i = C^{-1}H(H^\top C^{-1}H)^{-1}e_i.$$

These can be stacked up to write

$$A = \begin{bmatrix} a_0^\top \\ \vdots \\ a_{m-1}^\top \end{bmatrix} = \begin{bmatrix} e_0^\top (H^\top C^{-1}H)^{-1}H^\top C^{-1} \\ \vdots \\ e_{m-1}^\top (H^\top C^{-1}H)^{-1}H^\top C^{-1} \end{bmatrix} = (H^\top C^{-1}H)^{-1}H^\top C^{-1}$$

hence, the BLUE is

$$\boxed{\hat{\theta}_{\mathsf{BLUE}}(y) = \bar{A}y = (H^\top C^{-1}H)^{-1}H^\top C^{-1}y.}$$

This is indeed a linear estimator and it is easy to check that it is unbiased under our constraint that $\mathrm{E}[Y] = H\theta$. To confirm that it achieves the minimum variance, you would need to take the Hessian (see textbook).

BLUE Performance

The covariance of the BLUE for can be computed as

$$
\begin{aligned}
\mathrm{cov}[\hat{\theta}_{\mathsf{BLUE}}(Y)] &= \mathrm{E}\left\{(\hat{\theta}_{\mathsf{BLUE}}(Y)-\theta)(\hat{\theta}_{\mathsf{BLUE}}(Y)-\theta)^\top\right\} \\
&= \mathrm{E}\left\{(\bar{A}Y-\theta)(\bar{A}Y-\theta)^\top\right\} \\
&= \mathrm{E}\left\{(\bar{A}Y-\bar{A}H\theta)(\bar{A}Y-\bar{A}H\theta)^\top\right\} \\
&= \bar{A}\mathrm{E}\left\{(Y-H\theta)(Y-H\theta)^\top\right\}\bar{A}^\top \\
&= \bar{A}\mathrm{E}\left\{(Y-\mathrm{E}][Y])(Y-\mathrm{E}][Y])^\top\right\}\bar{A}^\top \\
&= \bar{A}C\bar{A}^\top \\
&= (H^\top C^{-1}H)^{-1}H^\top C^{-1}CC^{-1}H(H^\top C^{-1}H)^{-1} \\
&= (H^\top C^{-1}H)^{-1}.
\end{aligned}
$$

Hence

$$
\boxed{\mathrm{trace}\left[\mathrm{cov}[\hat{\theta}_{\mathsf{BLUE}}(Y)]\right] = \mathrm{trace}\left[(H^\top C^{-1}H)^{-1}\right].}
$$

Remarks

1. Calculation of the BLUE

$$\hat{\theta}_{\mathsf{BLUE}}(y) = \bar{A}y = (H^\top C^{-1} H)^{-1} H^\top C^{-1} y$$

does not require full knowledge of the joint pdf of the observations. All you need to know is
   - the covariance of the observations C and
   - how the mean of the observations relates to the unknown parameter, i.e. $\mathrm{E}[Y] = H\theta$.
2. This feature makes the BLUE particularly appealing in practical scenarios where the joint pdf of the observations may not be known, but the mean and covariance of the observations is known.
3. There may be significant performance loss, however, in using a linear estimator.

Example 2 revisited

Suppose we have random observations given by

$$Y_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,\beta) \qquad k = 0,\dots,n-1$$

and we wish to estimate the mean $\theta = \beta/2$. The MVU estimator is

$$\hat{\theta}_{\mathsf{MVU}}(y) = \frac{N+1}{2N}\max y$$

and its variance is

$$\mathrm{var}\left\{\hat{\theta}_{\mathsf{MVU}}(y)\right\} = \frac{\beta^2}{4N(N+2)}.$$

See Kay I Example 5.8 for the details.

Now let's compute the BLUE and see how it's performance compares...

Linear Model

If the observations can be written in the linear model form

$$Y = H\theta + W$$

where $H \in \mathbb{R}^{n \times m}$ is a known "mixing matrix" and $W \in \mathbb{R}^n$ is a zero-mean noise vector with covariance C (and otherwise arbitrary pdf), then

$$\hat{\theta}_{\mathsf{BLUE}}(y) = \bar{A}y = (H^\top C^{-1} H)^{-1} H^\top C^{-1} y$$

and

$$\mathrm{cov}[\hat{\theta}_{\mathsf{BLUE}}(Y)] = (H^\top C^{-1} H)^{-1}.$$

To see this, you just need to show that $\mathrm{E}[Y] = H\theta$ and $\mathrm{cov}[Y] = C$.

Note that this result holds irrespective of the pdf of W. The noise does not need to be Gaussian.

Linear Gaussian Model

If the observations can be written in the linear **Gaussian** model form

$$Y = H\theta + W$$

where $H \in \mathbb{R}^{n \times m}$ is a known "mixing matrix" and $W \in \mathbb{R}^n$ is distributed as $\mathcal{N}(0, C)$ then not only do the results on the previous slide still hold, but

$$\hat{\theta}_{\mathsf{BLUE}}(y) = \hat{\theta}_{\mathsf{MVU}}(y).$$

See Kay I: Theorem 4.1 (Minimum Variance Unbiased Estimator for the Linear Model) and Kay I: Section 4.5 (Extension to the Linear Model) for the derivation of $\hat{\theta}_{\mathsf{MVU}}(y)$.

Consequence: In this special case, there is no loss of performance when using the BLUE. The BLUE is also the MVU estimator.

Conclusions

- Read Kay I: Chapter 6 (especially check out the signal processing example in Section 6.6)
- Best Linear Unbiased Estimators are important **practical** estimators:
  - Can usually be computed even when the MVU estimator can't.
  - Doesn't require full knowledge of the joint pdf of the observations.
  - BLUE=MVU in the linear Gaussian model (assumed in lots of real-world applications)
  - Suitable for implementation on DSP or FPGA.
- BLUE is **not suitable** unless the mean of the observations is linear in the parameters, i.e. $\mathrm{E}[Y] = H\theta$. The whole derivation breaks down if this condition isn't true.
- It may be possible to **transform the observations** in some unsuitable cases, i.e. $Z = f(Y)$ where f is a nonlinear function, to make them suitable for BLUE such that $\mathrm{E}[Z] = H\theta$. See Kay I: Problem 6.5.
- A BLUE may perform significantly worse than an MVU estimator in some scenarios.