Properties of a Linear Unbiased Minimum Variance Estimator

Author: *R. Scott McIntire*

Date: Aug 1, 2023

Overview

A standard problem is to find the linear unbiased "minimum" variance estimator for the over determined problem†: $A\mathbf{x} = \mathbf{b}$, where A is nxm and $\mathbf{x}$ and $\mathbf{b}$ are $mx1$ and $nx1$ random variables respectively. The covariance of $\mathbf{b}$ is given as V and is assumed invertible. We look for a linear unbiased estimator for $\mathbf{x}$ of the form $\mathbf{x}^* = L\mathbf{b}$, where $L \in \mathbf{R}^{mn}$. Since $\mathbf{x}^*$ is unbiased we must have $E\mathbf{x} - \mathbf{x}^* = E[\mathbf{x} - L\mathbf{b}] = E[\mathbf{x} - LA\mathbf{x}]$. We satisfy this constraint by requiring that $LA = I$. This implies that A and L have rank m. The covariance of $L\mathbf{b}$ is LVL^T. The problem now is to find the best L such that $LA = I$ and the covariance matrix, LVL^T, is minimized in some sense.

The problem may be restated as choosing L to "minimize" the norm of LVL^T subject to the constraint $LA = I$. The question is what norm does one use. The "natural" norm of a matrix is $\sqrt{\mathrm{tr}\,(M^T M)}$. This follows from considering the standard basis in $\mathbf{R}^{nm}$ and using the usual norm of the vector space $\mathbf{R}^{nm}$. The matrix of interest, LVL^T is positive definite symmetric; in this case, each of the "trace" expressions, $\sqrt[k]{\mathrm{tr}\,(M^k)}$ $(k \in [1, \infty))$, is a norm for a positive definite matrix M_{mxm}.

Do the Norms Matter?

The question is, do different choices of norms lead to a different solution, L? To answer this, we solve the constrained minimization problem that results from each of these trace norms*:

$$F(L; k) = \mathrm{tr}\left((LVL^T)^k - (LA - I)\Lambda_k^T\right) \quad k \in [1, \infty)$$

where, $\Lambda_k \in \mathbf{R}^{mm}$ is a Lagrange multiplier.

The derivative of F with respect to L is given by:

$$DF(L)(H) = \mathrm{tr}\left[k(LVL^T)^{k-1}(HVL^T + LVH^T) - HA\Lambda_k^T\right]$$

At a minimum value, L^*, the derivative must vanish; therefore,

$$\mathrm{tr}\left[k(L^*VL^{*T})^{k-1}(HVL^{*T} + L^*VH^T) - HA\Lambda_k^T\right] = 0 \quad \forall H \in \mathbf{R}^{mn} \tag{1}$$

In the special case $k = 1$, equation (1) implies that $2VL^{*T} = A\Lambda_1^T$; or,

$$2L^{*T} = V^{-1}A\Lambda_1^T \tag{2}$$

---

† This means that $n > m$.
* Equivalently, we minimize the k^{th} power of the k^{th} trace norm.

Multiplying both sides by A^T and using the fact that $A^T L^{*T} = I$ yields $2I = A^T V^{-1} A \Lambda_1$. Since A has rank m, it follows that $\Lambda_1 = 2(A^T V^{-1} A)^{-1}$. Combining this with equation (2) gives

$$L^* = (A^T V^{-1} A)^{-1} A^T V^{-1}$$

We now proceed with the case when $k > 1$. Given any $\tilde{H} \in \mathbf{R}^{mm}$, let $H = \tilde{H} A^T V^{-1}$. Then equation (1) becomes:

$$\mathrm{tr}\left[k(L^* V L^{*T})^{k-1}(\tilde{H} A^T L^{*T} + L^* A \tilde{H}^T) - \tilde{H}(A^T V^{-1} A)\Lambda_k^T\right] = 0 \quad \forall \tilde{H} \in \mathbf{R}^{mm}$$

Or,

$$\mathrm{tr}\left[k(L^* V L^{*T})^{k-1}(\tilde{H} + \tilde{H}^T) - \tilde{H}(A^T V^{-1} A)\Lambda_k^T\right] = 0 \quad \forall \tilde{H} \in \mathbf{R}^{mm} \tag{3}$$

since $L^* A = I$ and $A^T L^{*T} = I$.

Given any $H' \in \mathbf{R}^{mm}$, consider $\tilde{H} = H'(A^T V^{-1} A)^{-1}$ with H' anti-symmetric. Therefore, $\tilde{H}$ is anti-symmetric; equation (3) becomes $\mathrm{tr}(H' \Lambda_k^T) = 0 \quad \forall H'_{mxm}$ with H' anti-symmetric. This implies that Λ_k is symmetric.

Equation (3) is true for all symmetric $\tilde{H}$. In this case it reads:

$$\mathrm{tr}\left[\left(2k(L^* V L^{*T})^{k-1} - \Lambda_k(A^T V^{-1} A)\right)\tilde{H}\right] = 0 \quad \forall \tilde{H} \in \mathbf{R}^{mm}\ ,\ \tilde{H} \text{ symmetric} \tag{3s}$$

Since $(L^* V L^{*T})^{k-1}$ and $\Lambda_k(A^T V^{-1} A)$ are both symmetric, equation (3s) implies that

$$2k(L^* V L^{*T})^{k-1} - \Lambda_k(A^T V^{-1} A) = 0$$

Since A has rank m, $(A^T V^{-1} A)^{-1}$ exists and we can solve for Λ_k.

$$\Lambda_k = 2k(L^* V L^{*T})^{k-1}(A^T V^{-1} A)^{-1}$$

Equation (1) now becomes:

$$\mathrm{tr}\left[k(L^* V L^{*T})^{k-1}(H V L^{*T} + L^* V H^T) - 2k H A (A^T V^{-1} A)^{-1}(L^* V L^{*T})^{k-1}\right] = 0 \quad \forall H \in \mathbf{R}^{mn}$$

Or,

$$\mathrm{tr}\left[\left((H V L^{*T} + L^* V H^T)/2 - H A (A^T V^{-1} A)^{-1}\right)(L^* V L^{*T})^{k-1}\right] = 0 \quad \forall H \in \mathbf{R}^{mn}$$

Which may be written

$$\mathrm{tr}\left[\left(H V L^{*T} + (L^* V H^T - H V L^{*T})/2 - H A (A^T V^{-1} A)^{-1}\right)(L^* V L^{*T})^{k-1}\right] = 0 \quad \forall H \in \mathbf{R}^{mn}$$

Since $(L^*VH^T - HVL^{*T})/2$ is anti-symmetric and $(L^*VL^{*T})^{k-1}$ is symmetric, the trace of their product is zero. Therefore we may write the last equation as

$$\operatorname{tr}\left[\left(HVL^{*T} - HA(A^TV^{-1}A)^{-1}\right)(L^*VL^{*T})^{k-1}\right] = 0 \quad \forall H \in \mathbf{R}^{mn}$$

Or,

$$\operatorname{tr}\left[H\left(VL^{*T} - A(A^TV^{-1}A)^{-1}\right)(L^*VL^{*T})^{k-1}\right] = 0 \quad \forall H \in \mathbf{R}^{mn}$$

This implies that $\left(VL^{*T} - A(A^TV^{-1}A)^{-1}\right)(L^*VL^{*T}) = 0$. Since L^* has rank m, $(L^*VL^{*T})^{-(k-1)}$ exists, so that $VL^{*T} - A(A^TV^{-1}A)^{-1} = 0$. Hence,

$$L^* = (A^TV^{-1}A)^{-1}A^TV^{-1}$$

Therefore, it does not matter which of the matrix norms we use; the answer is the same.

Related features of "the" best estimator

We now use the results of the trace norms above to show that the estimator obtained has the smallest maximum eigenvalue over all linear estimators.

Lemma 1. *If Ω is a closed set in R^m and if $x^* \in \Omega$ is a solution to the family of problems:* Min $\varphi_k(x)$ over $\{x \in \Omega : g(x) = 0\}$ $k \in [1, \infty)$*; further, if φ_k are functions which converge point-wise to the continuous function φ on Ω with the accretive property that $\lim_{\|x\| \to \infty} \varphi(x) \to \infty$, then x^* is a solution of* Min $\varphi(x)$ over $\{x \in \Omega : g(x) = 0\}$.

Proof:

Existence: We know there is a non empty set of points that satisfy $\{x \in \Omega : g(x) = 0\}$. If the minimum of φ is attained on this set, then a solution exists. Otherwise, the infimum on this set is finite, by the accretive property of φ. Let x_n be a sequence of points in this set such that $\varphi(x_n)$ converges to the infimum. Again, by the accretive property, the $\{x_n\}$ are bounded in Ω. Therefore, since Ω is closed, there exists a subsequence of the x_n which converges to a point, say $x' \in \Omega$. By the continuity of φ on Ω, x' must attain the infimum of φ on the set $\{x \in \Omega : g(x) = 0\}$; so that a solution exists.

We proceed to show that x^* is a solution. Let $\tilde{x}$ be a solution of the problem Min $\varphi(x)$ over $\{x \in \Omega : g(x) = 0\}$; then for any $\epsilon > 0$, we have, for sufficiently large n, (since φ_n converges point-wise to φ)

$$\varphi_n(\tilde{x}) < \varphi(\tilde{x}) + \epsilon$$

Since x^* is a solution for φ_n, we also have

$$\varphi_n(x^*) \leq \varphi_n(\tilde{x})$$

Combining the two gives

$$\varphi_n(x^*) \leq \varphi_n(\tilde{x}) < \varphi(\tilde{x}) + \epsilon$$

Or,

$$\varphi_n(x^*) \leq \varphi(\tilde{x}) + \epsilon$$

Taking the limit as $n \to \infty$ we have:

$$\varphi(x^*) \leq \varphi(\tilde{x}) + \epsilon$$

Since this is true for all $\epsilon > 0$ we necessarily have that

$$\varphi(x^*) \leq \varphi(\tilde{x})$$

Noting that x^* satisfies the constraint, $g(x) = 0$, we infer that x^* is also a solution.

Theorem 1. *The linear unbiased estimator previously obtained for the least squares problem, $L^*\mathbf{b} = (A^T V^{-1} A)^{-1} A^T V^{-1} \mathbf{b}$, is also the solution that minimizes the maximum eigenvalue of the estimator's covariance matrix, LVL^T, subject to the constraint $L_{mxn} A_{nxm} = I_{mxm}$.*

Proof: We apply Lemma 1 with Ω equal to the mxm symmetric matrices; φ equal to the L_2 matrix norm; and φ_n equal to the n^{th} trace norm. Theorem 1 follows after collecting the facts:

- The maximum eigenvalue of a symmetric matrix is the result of taking the limit of the trace norms of a symmetric matrix.
- The L_2 norm applied to a symmetric matrix is the maximum eigenvalue of that matrix.
- The L_2 norm is accretive.