Best Linear Unbiased Estimator

Report by Ethan Dowley

Introduction

In our lectures, we have encountered several Bayesian estimators such as the *posterior mean*,

$\hat{X} = \phi^*(Y)$, with $\phi^*: \mathbb{R}^{N_x} \to \mathbb{R}^{N_y}$ defined by $\phi^*(y) = \int_{\mathbb{R}^{N_x}} x \pi_X(x|y) dx$,

and the *maximum a posteriori* estimator (MAP). Here, we will briefly discuss another estimator which replaces the generally nonlinear conditional expectation (posterior mean).

Motivation

The minimum variance unbiased estimator (MVUE) is the optimal estimator for a given problem. However, finding this requires full knowledge of the probability density function (PDF) of the underlying process, which we rarely have in practise.

Instead, we restrict the estimator to be linear and seek the best linear unbiased estimator (BLUE). By definition, the BLUE is the unbiased linear estimator with the least variance. Unlike the MVUE, finding the BLUE only requires knowledge of the first two moments (mean and variance) of the PDF.

Model

Our standard forward model is $Y = h(X) + \varepsilon$, where we assume that the model errors ε are centred ($\mathbb{E}[\varepsilon] = 0$) and independent of X. No further assumptions about X or ε are required.

We postulate a linear estimator of the form $\hat{X} = AY + b$, where $A \in \mathbb{R}^{N_x \times N_y}$ and $b \in \mathbb{R}^{N_x}$ are determined by minimising the variance $\mathbb{E}\left[\left\|X - \hat{X}\right\|^2\right]$. (This is the definition of the BLUE.)

Derivation

Without taking the expectation value yet, we have

$$\left\|X - \hat{X}\right\|^2 = \big(X - (AY + b)\big)^T \big(X - (AY + b)\big)$$

$$= \sum_j \left[x_j^2 - x_j (AY + b)_j - (AY + b)_j x_j + (AY + b)_j (AY + b)_j\right]$$

$$= \sum_j \left[x_j^2 - x_j b_j - x_j \sum_i \ a_{ji} y_i - \sum_i \ a_{ji} y_i x_j - b_j x_j + \left(\sum_i \ a_{ji} y_i + b_j\right)\left(\sum_k \ a_{jk} y_k + b_j\right)\right]$$

$$= \sum_j \left[x_j^2 - 2 b_j x_j - 2 \sum_i \ x_j a_{ji} y_i + b_j^2 + 2 \sum_i \ a_{ji} y_i b_j + \sum_{i,k} a_{ji} a_{jk} y_i y_k\right]$$

We assume that all a_{ji} and b_j are independent variables. Since our goal is to minimise this expression, we set every partial derivative with respect to a_{ji} and b_j to zero:

$$\frac{\partial}{\partial a_{ji}} \left\|X - \hat{X}\right\|^2 = -2 x_j y_i + 2 b_j y_i + 2 \sum_k \ a_{jk} y_k y_i = 0, \ \ \forall\, j \in N_X, i \in N_Y$$

$$\frac{\partial}{\partial b_j} \left\|X - \hat{X}\right\|^2 = -2 x_j + 2 b_j + 2 \sum_i \ a_{ji} y_i = 0, \ \ \forall\, j \in N_X$$

Expressing these equations in matrix form and taking the expectation value, we have

$$-2\mathbb{E}[XY^T] + 2b\mathbb{E}[Y^T] + 2A\mathbb{E}[YY^T] = 0$$

$$-2\mathbb{E}[X] + 2b + 2A\mathbb{E}[Y] = 0$$

The second equation implies that $b = \bar{x} - A\bar{y}$, with $\bar{x} = \mathbb{E}[X]$ and $\bar{y} = \mathbb{E}[Y] = \mathbb{E}[h(X)]$. Substituting this into the first equation, dividing by 2 and using $\mathbb{E}[Y^T] = \bar{y}^T$, we get

$$0 = A\mathbb{E}[YY^T] - \mathbb{E}[XY^T] + (\bar{x} - A\bar{y})\mathbb{E}[Y^T] = A(\mathbb{E}[YY^T] - \bar{y}\bar{y}^T) - (\mathbb{E}[XY^T] - \bar{x}\bar{y}^T)$$

$= AP_{yy} - P_{xy}$, since by definition the cross-covariance matrix is

$$P_{xy} = \mathbb{E}[(X - \bar{x})(Y - \bar{y})^T] = \mathbb{E}[XY^T] - \bar{x}\bar{y}^T.$$

Therefore, $A = P_{xy}(P_{yy})^{-1}$ and $b = \bar{x} - A\bar{y} = \bar{x} - P_{xy}(P_{yy})^{-1}\bar{y}$.

Substituting these back into the original expression for the estimator gives

$\hat{X} = AY + b = P_{xy}(P_{yy})^{-1}Y + \bar{x} - P_{xy}(P_{yy})^{-1}\bar{y} = \bar{x} + P_{xy}(P_{yy})^{-1}(Y - \bar{y})$. This is the BLUE.

$$\boxed{\hat{X} = \bar{x} + P_{xy}(P_{yy})^{-1}(Y - \bar{y})}$$

Unbiased

We will briefly prove that the estimator we have found is indeed unbiased:

$$\mathbb{E}[\hat{X}] = \mathbb{E}\left[\bar{x} + P_{xy}(P_{yy})^{-1}(Y - \bar{y})\right] = \mathbb{E}[\bar{x}] + P_{xy}(P_{yy})^{-1}\mathbb{E}[Y - \bar{y}]$$

$= \bar{x} + P_{xy}(P_{yy})^{-1}\{\mathbb{E}[Y] - \mathbb{E}[\mathbb{E}[Y]]\} = \bar{x}$, which is the definition of an unbiased estimator.

Note that this proof did not depend on our assumption that the model errors are centred.

Kalman Mean

Since we assumed in our model $Y = h(X) + \varepsilon$ that X and ε were independent, it follows that $P_{yy} = P_{hh} + R$, where P_{hh} and R are the covariance matrices of $h(X)$ and ε respectively. So,

$$\hat{X} = \bar{x} + P_{xy}(P_{hh} + R)^{-1}(Y - \bar{y}).$$

If we now assume that the observation operator is linear, i.e. $h(X) = HX$ for some matrix $H \in \mathbb{R}^{N_y \times N_x}$, then $P_{xy} = P_{xx}H^T$, $P_{hh} = HP_{xx}H^T$ and $\bar{y} = H\bar{x}$. Using these values, we see that the BLUE reduces to the Kalman estimate for the mean:

$$\hat{X} = \bar{x} + P_{xy}(P_{hh} + R)^{-1}(Y - \bar{y}) = \bar{x} + P_{xx}H^T(HP_{xx}H^T + R)^{-1}(Y - H\bar{x})$$

(In the lectures, this expression was written as $\bar{x} - pH^T(HpH^T + R)^{-1}(H\bar{x} - Y)$, where $p = P_{xx}$ is the covariance matrix of X and the minus sign has been absorbed into the second pair of brackets.)

We originally derived the Kalman mean under the assumption that both X and ε were Gaussian, but here we see that those assumptions were unnecessary.

Gauss-Markov Theorem

"The ordinary least squares estimator (OLSE) has the lowest sampling variance within the class of linear unbiased estimators, if the errors in the linear regression model are uncorrelated, have equal variances and expectation value of zero." – Wikipedia

Let us assume that our model satisfies the conditions of the Gauss-Markov Theorem. It is easy to see the equivalence between the OLSE and the BLUE in this case. By definition, the BLUE is found by minimising the variance $\mathbb{E}\left[\|X - \hat{X}\|^2\right]$, while the OLSE is found by minimising the sum of squared residuals (SSR) $\sum_{n=1}^{N}\|X^n - \hat{X}^n\|^2$, where X^n is the value of X at time n.

Under these assumptions, $\mathbb{E}\left[\left\|X-\hat{X}\right\|^2\right]=\frac{\sum_{n=1}^{N}\|X^n-\hat{X}^n\|^2}{N}$.

Clearly, minimising the SSR (for the OLSE) is equivalent to minimising the variance (for the BLUE). The only difference is in the framing of the problem. However, this will not necessarily be the case for a model that does not satisfy these conditions.

Conclusion

We have used the framework introduced in the lectures to derive an expression for the BLUE, expanding on the explanation provided in the textbook by Reich and Cotter, and discussed two special cases where the BLUE reduces to the more familiar Kalman estimate or the OLSE.

References

[1] *Probabilistic Forecasting and Bayesian Data Assimilation* by Sebastian Reich and Colin Cotter

[2] BLUE estimator - GaussianWaves

[3] OLS Regression, Gauss-Markov, BLUE, and understanding the math | by Andrew Rothman | Towards Data Science

[4] Cross-covariance matrix - Wikipedia

[5] The Gauss-Markov Theorem and BLUE OLS Coefficient Estimates - Statistics By Jim

[6] 7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression - Statistics By Jim

[7] Gauss–Markov theorem - Wikipedia

[8] Chapter 1 Introduction (sunysb.edu)