Generalized Least Squares

Hakbae Lee

The Department of Statistics and Data Science, Yonsei University

Consider a full rank parameterization

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad \text{with } \mathrm{E}(\epsilon) = 0 \text{ and } \mathrm{Cov}(\epsilon) = \sigma^2\Sigma > 0.$$

By Spectral Decomposition of Σ,

$$\Sigma = \Gamma^T \Lambda \Gamma = \Gamma^T \Lambda^{1/2}\Lambda^{1/2}\Gamma = \Gamma^T \Lambda^{1/2}\Gamma^T\Gamma\Lambda^{1/2}\Gamma = \Sigma^{1/2}\Sigma^{1/2}$$

$$\mathbf{Z} \equiv \Sigma^{-1/2}\mathbf{Y} = \Sigma^{-1/2}\mathbf{X}\beta + \Sigma^{-1/2}\epsilon = \mathbf{W}\beta + \epsilon^*$$

and

$$\begin{aligned}
\hat{\beta} &= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z} = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}\mathbf{Y} \\
\mathrm{E}(\hat{\beta}) &= (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}\mathbf{X}\beta = \beta \\
\mathrm{Cov}(\hat{\beta}) &= \sigma^2(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1} \\
\hat{\sigma}^2 &= \frac{||Z - \mu_Z||^2}{n-p} = \frac{(\mathbf{Y} - \hat{\mu})^T\Sigma^{-1}(\mathbf{Y} - \hat{\mu})}{n-p}
\end{aligned}$$

Generalized Least Squares

- The projection matrix is $\Sigma^{-1/2}\mathbf{X}(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1/2}$, which is symmetric, and hence is an orthogonal projection.
- Now all computations have been done in the z coordinates, so in particular $x^T\beta$ estimates $\mu_z = \Sigma^{-1/2}\mu$.
- Since linear combinations of Gauss-Markov estimates are Gauss-Markov, it follows immediately that

$$\hat{\mu}_z = \Sigma^{-1/2}\hat{\mu}$$

A direct solution via inner products

- We can approach the problem of determining the generalized least squares estimators in a different way by viewing Σ as determining an inner product.
- We do this by returning to first principles, carefully defining means and covariances in a general inner product space.
- Let $x, y \in \mathbf{R}^n$, and $(x, y) = x^T y$ be the usual inner product.
- Choose a basis $\{e_1, \ldots, e_n\}$, the usual coordinate vectors. Then a random vector x has coordinates $(e_i, x) = x_i$.

A direct solution via inner products

Definition 1. $\mathrm{E}(x) = \mu = (\mu_i)$ where $\mu_i = \mathrm{E}(e_i, x)$. For any $a \in \mathbf{R}^n$,

$$\begin{aligned} \mathrm{E}((a, x)) &= \mathrm{E}\left(\left(\sum_{i=1}^{n} a_i e_i, x\right)\right) \\ &= \mathrm{E}\left(\sum_{i=1}^{n} a_i (e_i, x)\right) \\ &= \sum_{i=1}^{n} a_i \mu_i = (a, \mu) \end{aligned}$$

Thus, another characterization of μ is: μ is the unique vector that satisfies $\mathrm{E}((a, x)) = (a, \mu)$ for all $a \in \mathbf{R}^n$.

A direct solution via inner products

Now, turn to covariances. Use the same set-up as above. If $E(x_i^2) < \infty$, then

$$\mathrm{Cov}(x_i, x_j) = (x_i - \mu_i)(x_j - \mu_j) = \sigma_{ij} = \sigma_{ji}$$

exists for all i, j, and defines $\Sigma = (\sigma_{ij})$. For any $a, b \in \mathbf{R}^n$,

$$\begin{aligned}
\mathrm{Cov}((a, x), (b, x)) &= \mathrm{Cov}\left(\left(\sum_{i=1}^{n} a_i x_i, \sum_{j=1}^{n} b_j x_j\right)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}(x_i, x_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \sigma_{ij} = (a, \Sigma b)
\end{aligned}$$

A direct solution via inner products

Definition 2 Assume $E((a,x)^2) < \infty$. The unique non-negative definite linear transformation $\Sigma : V \to V$ that satisfies $\mathrm{Cov}((a,x),(b,x)) = (a, \Sigma b)$ for all $a, b \in V$ is called the covariance of X and is denoted $\mathrm{Cov}(x)$.

Theorem 1 Let $Y \in V$ with inner product $(\bullet,\bullet), \mathrm{Cov}(Y) = \Sigma$. Define another inner product $(\bullet,\bullet)$ on V by $[x,y] = (x, Ay)$ for some positive definite A. Then the covariance of X in the inner product space $(V,[\bullet,\bullet])$ is ΣA

Note 1: This shows that if $\mathrm{Cov}(X)$ exists in one inner product, it exists in all inner products. If $\mathrm{Cov}(X) = \Sigma$ in $(V,(\bullet,\bullet))$, then if $\Sigma > 0$ in the inner product $[x,y] = (x, \Sigma^{-1} y)$, the covariance is $\Sigma^{-1}\Sigma = I$.

A direct solution via inner products

Theorem 2 Suppose $\text{Cov}(X) = \Sigma$ in $(V, (\bullet, \bullet))$. If Σ_1 is symmetric on $(V, (\bullet, \bullet))$, and $Cov((a, x)) = (a, \Sigma_1 a)$ for all $a \in V$, then $\Sigma_1 = \Sigma$. This implies that the covariance is unique.

Consider the inner product space given by $(\mathbf{R}^n, [\bullet, \bullet])$, where $[x, y] = (x, \Sigma^{-1} y)$, $\text{E}(Y) = \mu \in \mathcal{E}$ and $\text{Cov}(Y) = \sigma^2 \Sigma$.

Let P_Σ be the projection on $\mathcal{E}$ in this inner product space, and let $Q_\Sigma = I - P_\Sigma$, so $y = P_\Sigma y + Q_\Sigma y$.

A direct solution via inner products

Theorem 3 With $[x, y] = (x, \Sigma^{-1}y)$,

$$P_\Sigma = \mathbf{X}(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}$$

is an orthogonal projection.

Theorem 4 The OLS estimate $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ and the GLS estimate $\tilde{\beta} = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}\mathbf{Y}$ are the same if and only if

$$\mathcal{C}(\Sigma^{-1}\mathbf{X}) = \mathcal{C}(\mathbf{X}).$$

Corolory 1

$$\mathcal{C}(\Sigma^{-1}\mathbf{X}) = \mathcal{C}(\mathbf{X}) = \mathcal{C}(\Sigma\mathbf{X}).$$

So Σ need not be inverted to apply the theory.

A direct solution via inner products

- To use this equivalence theorem (due to W. Kruskal), we usually characterize the Σs for a given X for which $\hat{\beta} = \tilde{\beta}$.
- If X is completely arbitrary, then only $\Sigma = \sigma^2 I$ works.
- **Intra-class correlation model:** Let $J_n \in \mathcal{C}(\mathbf{X})$. Then any Σ of the form
$$\Sigma = \sigma^2(1-\rho)I + \sigma^2 \rho J_n J_n^T$$
with $-1/(n-1) < \rho < 1$ will work.
- To apply the theorem, we write,
$$\Sigma \mathbf{X} = \sigma^2(1-\rho)\mathbf{X} + \sigma^2 \rho J_n J_n^T \mathbf{X}$$
so for $i > 1$, the i-th column of $\Sigma \mathbf{X}$ is
$$(\Sigma \mathbf{X})_i = \sigma^2(1-\rho)X_i + \sigma^2 \rho J_n a_i$$
with $a_i = J_n^T \mathbf{X}$.

A direct solution via inner products

- Thus, the i-th column of $\Sigma\mathbf{X}$ is a linear combination of the i-th column of $\mathbf{X}$ and the column of 1s.
- For the first column of $\Sigma\mathbf{X}$, we compute $a_1 = J_n$ and

$$(\Sigma\mathbf{X})_1 = \sigma^2(1-\rho)J_n + n\sigma^2\rho J_n = \sigma^2(1+\rho(n-1))J_n$$

so $\mathcal{C}(\Sigma\mathbf{X}) = \mathcal{C}(\mathbf{X})$ as required, provided that $1+\rho(n-1) \neq 0$ or $\rho > -1/(n-1)$.