MATH 531 Homework 4
Generalized least squares and variance estimation
February 7, 2025

Consider a linear model for data as

$$\mathbf{y} = X\beta + \mathbf{e}$$

where $\mathbf{y}$ is an vector of length n and X is a full rank, $n \times p$ matrix. Assume that $cov(\mathbf{e}) = \Sigma$ ($n \times n$ is known only up to a constant. That is we are given a matrix, Ω, that has full rank, where $\Sigma = \sigma^2\Omega$ but we don't know the scalar parameter σ^2. Assume throughout that Ω, being proportional to the covariance matrix for the data, is positive definite.

Although this may seem a strange setup it is exactly how we think about the OLS model : $\Sigma = \sigma^2 I_n$ with I_n the identity matrix.

Now suppose you create the "star" model. $\mathbf{y}^* = \Omega^{-1/2}\mathbf{y}$, $X^* = \Omega^{-1/2}X$, and $\mathbf{e}^* = \Omega^{-1/2}\mathbf{e}$. Here $\Omega^{-1/2}$ is the symmetric version of the square root obtained by the eigen decomposition for Ω. Symmetric version so we don't have to worry about keeping track of transposes.

$$\mathbf{y}^* = X^*\beta + \mathbf{e}^*$$

This is motivated by simply multiplying both sides of the OLS model by $\Omega^{-1/2}$.

Note that in a (practical) data analysis both $\mathbf{y}^*$ and X^* can be directly computed from the data without estimating any unknown parameters. (However, often Ω will depend on other parameters that we don't know – but let's not go there!)

1. (a) Explain how to construct $\Omega^{-1/2}$ based on an eigen decomposition of Ω.

   Let $\Omega = Q\Lambda Q^T$ be the eigendecomposition of Ω, where $Q \in \mathbb{R}^{n\times n}$. Note that we use Q^T instead of Q^{-1} because Ω is a real-valued symmetric matrix. Furthermore, let $f(A) = A^{-\frac{1}{2}}$. Since Λ is a diagonal matrix, we can apply f to the eigendecomposition as such:

   $$\begin{aligned} f(\Omega) &= Qf(\Lambda)Q^T \\ &= Q\Lambda^{-\frac{1}{2}}Q^T \end{aligned}$$

   We know that Ω is positive definite, so we can say that $\Omega^{-\frac{1}{2}}$ is a symmetric positive definite matrix as well.

(b) Show that $E(\mathbf{e}^*) = 0$, $E(\mathbf{y}^*) = X^*\beta$ and $COV(\mathbf{e}^*) = COV(\mathbf{y}^*) = \sigma^2 I_n$

Proof. For this, we use the definitions given to us by the "star" model. First, start with $\mathbb{E}[e^*]$. We know that $\Omega^{-\frac{1}{2}}$ is not a random matrix, so we can extract it from the expectation expression. Hence, we have

$$\begin{aligned}\mathbb{E}[e^*] &= \mathbb{E}[\Omega^{-\frac{1}{2}}e] \\ &= \Omega^{-\frac{1}{2}}\mathbb{E}[e] \\ &= \Omega^{-\frac{1}{2}} \cdot 0 \\ &= 0\end{aligned}$$

Next, we derive $\mathbb{E}[y^*]$ using the definition of y^* as stated above.

$$\begin{aligned}\mathbb{E}[y^*] &= \mathbb{E}[\Omega^{-\frac{1}{2}}y] \\ &= \Omega^{-\frac{1}{2}}\mathbb{E}[y] \\ &= \Omega^{-\frac{1}{2}}X\beta\end{aligned}$$

We know that $X^* = \Omega^{-\frac{1}{2}}X$. Therefore, we have

$$\mathbb{E}[y^*] = \Omega^{-\frac{1}{2}}X\beta = X^*\beta$$

Lastly, we want to show that $\text{Cov}[e^*] = \text{Cov}[y^*] = \sigma^2 I_n$. First, we start with $\text{Cov}[e^*]$.

$$\begin{aligned}\text{Cov}[e^*] &= \text{Cov}[e^*, e^*] \\ &= \text{Cov}[\Omega^{-\frac{1}{2}}e^*, \Omega^{-\frac{1}{2}}e^*] \\ &= \Omega^{-\frac{1}{2}}\text{Cov}[e]\Omega^{-\frac{1}{2}}\end{aligned}$$

Note that $\Omega^{-\frac{1}{2}}$ is symmetric, so it is equal to its transpose. Above, the covariance of e is defined as $\text{Cov}[e] = \sigma^2\Omega$, so we can substitute this back into the expression above.

$$\Omega^{-\frac{1}{2}}\text{Cov}[e]\Omega^{-\frac{1}{2}} = \sigma^2\Omega^{-\frac{1}{2}}\Omega\Omega^{-\frac{1}{2}}$$

We can expand this out and simplify.

$$\sigma^2\Omega^{-\frac{1}{2}}\Omega\Omega^{-\frac{1}{2}} = \sigma^2 Q\Lambda^{-\frac{1}{2}}Q^T Q\Lambda Q^T Q\Lambda^{-\frac{1}{2}}Q^T$$

Q is orthogonal, so we have

$$\sigma^2\Omega^{-\frac{1}{2}}\Omega\Omega^{-\frac{1}{2}} = \sigma^2 Q\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}Q^T$$

The product $\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}$ is an elementwise operation since each matrix is a diagonal matrix. Since $\Lambda^{-\frac{1}{2}}$ is the inverse square root of Λ, then the expression simply becomes

$$\sigma^2 Q\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}Q^T = \sigma^2 Q I_n Q^T = \sigma^2 I_n$$

Thus, $\text{Cov}[e^*] = \sigma^2 I_n$. We can apply the same procedure to $\text{Cov}[y^*]$.

$$\text{Cov}[y^*] = \Omega^{-\frac{1}{2}}\text{Cov}[y]\Omega^{-\frac{1}{2}}$$

We know that $\text{Cov}[y] = \text{Cov}[e] = \sigma^2\Omega$, so the expression above becomes

$$\begin{aligned}\text{Cov}[y^*] &= \sigma^2\Omega^{-\frac{1}{2}}\Omega\Omega^{-\frac{1}{2}} \\ &= \sigma^2 I_n\end{aligned}$$

Thus, $\text{Cov}[y^*] = \text{Cov}[e^*] = \sigma^2 I_n$. $\square$

(c) Show that if $\mathbf{c}^T\beta$ is estimable in the original model it is also estimable in the "star" model.

Proof. We know that if $c^T\beta$ is estimable, then there exists some $a \in \mathbb{R}^n$ such that

$$c^T\beta = \mathbb{E}[a^T y].$$

Now, consider the transformed "star" model, where we substitute $y^* = \Omega^{-1/2}y$. Then, we have:

$$\begin{aligned}c^T\beta &= \mathbb{E}[a^T y^*] \\ &= \mathbb{E}[a^T\Omega^{1/2}y] \\ &= a^T\mathbb{E}[\Omega^{1/2}y] \\ &= a^T\Omega^{-1/2}X\beta \\ &= a^T X^*\beta.\end{aligned}$$

Since $a^T X^*\beta$ follows the same form as in the original model, this shows that $c^T\beta$ remains estimable in the "star" model if it was estimable in the original model. $\square$

2. (a) Let $\hat{\beta}^*$ be the OLS estimate based on the "star" model. Use the Gauss-Markov theorem to argue the optimality of this estimate.

   Previously, we proved that the "star" model is both strictly exogenous and the errors are homoskedastic. We consider the estimator $\hat{\beta}^* = (X^{*T}X^*)^{-1}X^{*T}y^*$. We can argue that $\widehat{\beta}^*$, the OLS estimate of the star model, is the BLUE for the star model via the Gauss-Markov Theorem. We know that $\lambda^{*T}\beta$ is estimable, so it follows that the OLS estimate for $\lambda^{*T}\beta$, or $\widehat{\beta}^*$, is the BLUE.

   (b) Explain how to use the star model and some moment conditions to guarantee an unbiased and minimum variance estimate for σ^2.
   BTW: Will the square root of your estimate also be unbiased for σ (YES,NO) ?)

   We defined $\widehat{\beta}^*$ above. So we use this to estimate $y^* = X^*\widehat{\beta}^* + r^*$, where $r*$ denotes the residuals of the model. We can rearrange to find that $r^* = y^* - X^*\widehat{\beta}^*$. The unbiased estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{RSS}{n-p} = \frac{\|r^*\|^2}{n-p}$$

   where p is the number of columns in X^* and n is the number of observations. Under normal OLS assumptions, this is an unbiased and minimal-variance estimator of σ^2.

   We know that if $\hat{\sigma}^2$ is an unbiased estimator, then $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$. However, since taking the square root is a nonlinear operation, then it follows that $\mathbb{E}\left[\sqrt{\hat{\sigma}^2}\right] \neq \sqrt{\mathbb{E}[\hat{\sigma}^2]}$. Therefore, the square root of the estimate is not unbiased.

3. Writing the results in terms of the original data – your previous answers are in the form of the "star" variables $\mathbf{y}^*$ and X^*.

   (a) Express your estimate of β in terms of the original $\mathbf{y}$ and $\mathbf{X}$.

   We have from earlier in (2a) that $\hat{\beta}^* = (X^{*T}X^*)^{-1}X^{*T}y^*$. We can use the definitions given to us by the star model to expand in terms of y and X from the original model.

$$\begin{aligned} \hat{\beta}^* &= (X^{*T}X^*)^{-1}X^{*T}y^* \\ &= ((\Omega^{-\frac{1}{2}}X)^T\Omega^{-\frac{1}{2}}X)^{-1}(\Omega^{-\frac{1}{2}}X)^T\Omega^{-\frac{1}{2}}y \\ &= (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}y \end{aligned}$$

   (b) Express your estimate of σ^2 in terms of the original data and covariates.

We take our definition of $\hat{\sigma}^2$ from (2b) and expand to express it in terms of original data.

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\|e^*\|^2}{n-p} \\ &= \frac{e^{*T}e^*}{n-p} \\ &= \frac{(\Omega^{-\frac{1}{2}}e)^T\Omega^{-\frac{1}{2}}e}{n-p} \\ &= \frac{e^T\Omega^{-1}e}{n-p}\end{aligned}$$

(c) Is your estimate for $X\hat{\beta}$ based on the projection of $\mathbf{y}$ on the column space of X? (YES,NO).

No. Instead of a standard orthogonal projection, it is a weighted projection with respect to Ω^{-1}.

4. This problem explores the optimality of the estimate of the variance motivated by least squares fitting. The idea is to use Monte Carlo simulation in R and we will focus on just a simple random sample instead of a linear model to make things easier.

(a) Let $\{y_i\}$ be a random sample from a $N(\mu, \sigma^2)$. Of course from intro stats we know that μ is estimated by the sample mean and σ^2 estimated by the sample variance. We also know that $\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$ will have a Chi-squared distribution with $n-1$ degrees of freedom. (We will derive more general results later in the course.)

- Explain how we can also find these estimates using a simple linear regression model.

We can estimate μ and σ^2 with a linear model via least squares estimation. We can estimate μ as the intercept of the model. We can estimate σ^2 by calculating the RSS and dividing it by the degrees of freedom.

- Report the variance of $\hat{\sigma}^2$

The variance of $\hat{\sigma}^2$ is $\text{Var}[\hat{\sigma}^2] = \frac{2\sigma^4}{n}$. This is a result from standard introductory statistics. I found the proof here.

(b) Now consider the estimate for σ based on least absolute deviations.

$$\hat{\gamma} = (1/(nC)) \sum_{i=1}^{n} |y_i - \bar{y}|$$

and estimate σ^2 by squaring this: $\hat{\gamma}^2$.

- Derive the constant C to make this closer to unbaised using $C = E(|Z|)$ where $Z \sim N(0,1)$.

(I think it is $2/\sqrt{2\pi}$ – how would you check this?)

Normal distributions are symmetric, so we have

$$\mathbb{E}[|Z|] = 2\int_{-\infty}^{\infty} z \cdot f(z)dz$$

where $f(z) = \frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}$ is the standard probability density function. If we use $u-$substitution, where $u = \frac{z^2}{2}$ and $du = zdz$, when z goes from $0 \to \infty$, we have

$$2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u}du = \frac{2}{\sqrt{2\pi}}.$$

- Generate 10^5 random samples of size n=30, from a $N(10,(2)^2)$. For each sample compute the estimate $\hat{\gamma}$ from the Monte Carlo results estimate the bias, variance and root mean squared error of the estimator $\hat{\gamma}^2$ for σ^2. Also find the usual one – $\hat{\sigma}^2$ – the sample variance.

Code:

```
    n_sims <- 10e5
n <- 30

# True parameters
mu <- 10
sigma_true <- 2
sigma2_true <- sigma_true^2

# Derived Value of C
C <- 2/sqrt(2*pi)


gamma_sq_vals <- numeric(n_sims)
sample_var_vals <- numeric(n_sims)

# Main Monte Carlo loop
for (i in 1:n_sims) {

  # Generate a random sample of n=30 from Normal(10, 2^2)
  y <- rnorm(n, mean = mu, sd = sigma_true)

  # LAD-based estimate: hat(gamma)
  ybar <- mean(y)
```

```r
  gamma_hat <- (1/(n*C)) * sum(abs(y - ybar))
  gamma_sq_vals[i] <- gamma_hat^2

  # Usual sample variance: hat(sigma^2)
  sample_var_vals[i] <- var(y)
}


# 1) For gamma^2
mean_gamma_sq <- mean(gamma_sq_vals)
bias_gamma_sq <- mean_gamma_sq - sigma2_true
var_gamma_sq <- var(gamma_sq_vals)
rmse_gamma_sq <- sqrt(mean((gamma_sq_vals - sigma2_true)^2))

# For sigma hat^2
mean_s_var <- mean(sample_var_vals)
bias_s_var <- mean_s_var - sigma2_true
var_s_var <- var(sample_var_vals)
rmse_s_var <- sqrt(mean((sample_var_vals - sigma2_true)^2))

cat("(hat(gamma)^2) \n")
cat("Bias:     ", bias_gamma_sq, "\n")
cat("Variance: ", var_gamma_sq,  "\n")
cat("RMSE:     ", rmse_gamma_sq, "\n\n")

cat("(hat(sigma^2)) \n")
cat("Bias:     ", bias_s_var, "\n")
cat("Variance: ", var_s_var,  "\n")
cat("RMSE:     ", rmse_s_var, "\n")
```

Output:

```
(hat(gamma)^2)
Bias:      -0.05799057
Variance:  1.211652
RMSE:      1.102276

(hat(sigma^2))
Bias:      -7.190902e-05
Variance:  1.102798
RMSE:      1.050141
```

- Compare your results to the estimator $\hat{\sigma}^2$ – which is better according to these metrics?

The standard sample variance estimator seems to outperform the $\hat{\gamma}^2$ estimator for bias, variance, and RMSE.

(c) Compare your results using both $\hat{\gamma}$ and $\hat{\sigma}$ as estimators for σ.

The sample variance estimator is unbiased, as it is negligibly small, whereas the LAD estimator has slight negative bias. The variance of the sample variance estimator is smaller than the LAD estimator. The RMSE of the sample variance estimator is less than the LAD estimator. Overall, the sample variance-based estimator $\hat{\sigma}^2$ is preferred due to lower bias, lower variance, and lower RMSE.

(d) Under what circumstances might you prefer to use $\hat{\gamma}$ as an estimate for σ?

A big reason to use the $\hat{\gamma}$ estimator is when the data does not fit well to a normal distribution, as it is more robust to data that has outliers or non-normal distributions.