2013 Qualifying Exam Section 2

February 21, 2019

1 Question 1

2 Question 2

Consider the linear model

$$Y = X\beta + Z\gamma + \epsilon,$$

where Y is $n \times 1$, X is $n \times p$ of rank p, Z is $n \times q$ of rank q, β is an unknown $p \times 1$ parameter vector, γ is $q \times 1$, $\epsilon \sim N_n(0, R)$, $\sim N_q(0, D)$, R and D are positive definite matrices, ϵ and γ are independent, and $N_n(a, b)$ is an n variate normal random variable with mean vector a and covariance matrix b.

2.a

For known R and D, the distribution of $Y|X, \gamma \sim N(X\beta + Z\gamma, R)$. Derive the marginal distribution of $Y|X$.

Solution:
Note that

$$\mathbb{E}Y|X = \mathbb{E}\left\{\mathbb{E}Y|X, \gamma\right\} = \mathbb{E}(X\beta + Z\gamma) = X\beta$$

$$\begin{aligned}
\text{Cov}(Y|X) &= \text{Cov}\left\{\mathbb{E}Y|X, \gamma\right\} + \mathbb{E}\text{Cov}\left\{Y|X, \gamma\right\} \\
&= \text{Cov}\left\{X\beta + Z\gamma\right\} + \mathbb{E}(R) \\
&= Z\text{Cov}(\gamma)Z^T + R \\
&= ZDZ^T + R
\end{aligned}$$

Since the overall distribution is multivariate normal, so are the conditional distributions. Thus, it follows that $Y|X \sim N(X\beta, ZDZ' + R)$

2.b

In the following, continue to assume that R and D are known and treat γ as an unknown parameter in $Y|X, \gamma$.

2.b.1

Show that the predictor of γ given by $\hat{\gamma} = DZ'V^{-1}(Y - X\hat{\beta})$ satisfies the conditional likelihood equations for (β, γ), where $\hat{\beta}$ is the MLE for β and $V = ZDZ' + R$

Solution: The pdf of $\gamma|y, X$ can be found by:

$$\begin{aligned} p(\gamma|y,X) &\propto p(y|\gamma,X)p(\gamma|X) \\ &\propto \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)^T R^{-1}(y - X\beta - Z\gamma)\right\}\exp\left\{-\frac{1}{2}\gamma^T D^{-1}\gamma\right\} \\ &\propto \exp\left\{-\frac{1}{2}(y - X\beta)^T R^{-1}(y - X\beta) - \frac{1}{2}\gamma^T(Z^T R^{-1}Z + D^{-1})\gamma - 2\gamma^T Z^T R^{-1}(y - X\beta)\right\} \end{aligned}$$

Thus, we can write the log likelihood as

$$\ell = -\frac{1}{2}(y - X\beta)^T R^{-1}(y - X\beta) - \frac{1}{2}\gamma^T(Z^T R^{-1}Z + D^{-1})\gamma + \gamma^T Z^T R^{-1}(y - X\beta)$$

Note that we can write the sum of squares as

$$y^T R^{-1}y + \beta^T X^T R^{-1}X\beta - 2\beta^T X^T R^{-1}(y - Z\gamma)$$

The score equations are

$$\begin{aligned} S_n^{(1)}(\beta,\gamma) &= \frac{\partial \ell}{\partial \beta} = -(X^T R^{-1}X\beta - X^T R^{-1}(y - Z\gamma)) \\ S_n^{(2)}(\beta,\gamma) &= \frac{\partial \ell}{\partial \gamma} = -(Z^T R^{-1}Z + D^{-1})\gamma + Z^T R^{-1}(y - X\beta) \end{aligned}$$

We want to show $\hat{\gamma}$ satisfies the conditional likelihood equations, i.e., that $S_n(\hat{\beta}, \hat{\gamma}) = 0$. Thus, we have

$$
\begin{aligned}
S_n^{(2)}(\hat{\beta}, \hat{\gamma}) &= -(Z^T R^{-1} Z + D^{-1})(D Z^T V^{-1}(y - X\hat{\beta})) + Z^T R^{1}(y - X\hat{\beta}) \\
&= -(Z^T R^{-1} Z + D^{-1})(D Z^T V^{-1}(y - X\hat{\beta})) - Z^T R^{-1}(y - X\hat{\beta}) \\
&= -Z^T R^{-1} Z D Z^T V^{-1}(y - X\hat{\beta}) - D^{-1} D Z^T V^{-1}(y - X\hat{\beta}) + Z^T R^{-1}(y - X\hat{\beta}) \\
&= -Z^T \left[R^{-1} Z D Z^T V^{-1} + V^{-1} - R^{-1}\right](y - X\hat{\beta}) \\
&= -Z^T \left[R^{-1}(V - R) V^{-1} + V^{-1} - R^{-1}\right](y - X\hat{\beta}) \\
&= -Z^T \left[R^{-1} V V^{-1} - R^{-1} R V^{-1} + V^{-1} - R^{-1}\right](y - X\hat{\beta}) \\
&= -Z^T \left[R^{-1} - V^{-1} + V^{-1} - R^{-1}\right](y - X\hat{\beta}) \\
&= 0
\end{aligned}
$$

Thus, $(\hat{\beta}, \hat{\gamma})$ satisfy the conditional likelihood equations since $\hat{\beta}$ is such that $S_n^{(1)}(\hat{\beta}, \hat{\gamma}) = 0$.

2.b.2

Derive the exact distribution of $\hat{\gamma}$

Solution:
Setting the first score equation to 0, we have

$$\begin{aligned}
S_n^{(1)}(\beta, \hat{\gamma}) &\overset{\text{SET}}{=} 0 \\
&\implies X^T R^{-1} X\hat{\beta} - 2X^T R^{-1}(y - Z\gamma) = 0 \\
&\implies X^T R^{-1}[X\hat{\beta} + ZDZ^T V^{-1}(y - X\hat{\beta})] = X^T R^{-1} y \\
&\implies X^T R^{-1}[I - ZDZ^T V^{-1}]X\hat{\beta} = X^T R^{-1}[I - ZDZ^T V^{-1}]y \\
&\implies X^T R^{-1}[I - (V - R)V^{-1}]X\hat{\beta} = X^T R^{-1}[I - (V - R)V^{-1}]y \\
&\implies X^T R^{-1}[I - (I - RV^{-1})]X\hat{\beta} = X^T R^{-1}[I - (I - RV^{-1})]y \\
&\implies X^T R^{-1} R V^{-1} X\hat{\beta} = X^T R^{-1} R V^{-1} y \\
&\implies X^T V^{-1} X\hat{\beta} = X^T V^{-1} y \\
&\implies \hat{\beta} = [X^T V^{-1} X]^{-1} X^T V^{-1} y
\end{aligned}$$

Thus,

$$\begin{aligned}
\hat{\gamma} &= DZ^T V^{-1}(Y - X\hat{\beta}) \\
&= DZ^T V^{-1}[Y - (X^T V^{-1} X)^{-1} X^T V^{-1} Y] \\
&= DZ^T V^{-1}[I - (X^T V^{-1} X)^{-1} X^T V^{-1}]Y
\end{aligned}$$

Since y is multivariate normal and we have a linear form in y, we have

$$\hat{\gamma} \sim N(\mu_\gamma, \Sigma_\gamma)$$

where $\mu_\gamma = \mathbb{E}\hat{\gamma}$ and $\Sigma_\gamma = \text{Cov}(\hat{\gamma})$

$$\begin{aligned}
\mathbb{E}\hat{\beta} &= \mathbb{E}(X^T V^{-1} X)^{-1} X^T V^{-1} y \\
&= (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbb{E} y \\
&= (X^T V^{-1} X)^{-1} X^T V^{-1} X\beta \\
&= \beta
\end{aligned}$$

Thus,

$$\begin{aligned}
\mu_\gamma &= \mathbb{E}\hat{\gamma} \\
&= \mathbb{E} DZ^T V^{-1}(Y - X\hat{\beta}) \\
&= DZ^T V^{-1}(\mathbb{E}Y - X\mathbb{E}\hat{\beta}) \\
&= DZ^T V^{-1}(X\beta - X\beta) \\
&= 0
\end{aligned}$$

Now, note that

$$
\begin{aligned}
y - X\hat{\beta} &= y - X(X^T V^{-1} X)^{-1} X^T V^{-1} y \\
&= [I - (X^T V^{-1} X)^{-1} X^T V^{-1}] y
\end{aligned}
$$

Moreover

$$
\begin{aligned}
A &\equiv X(X^T V^{-1} X)^{-1} X^T V^{-1} \\
&= V^{1/2} V^{-1/2} X (X^T V^{-1/2} V^{-1/2} X)^{-1} X^T V^{-1/2} V^{-1/2} \\
&= V^{1/2} B (B^T B)^{-1} B^T V^{-1/2} \\
&= V^{1/2} M V^{-1/2}
\end{aligned}
$$

where $B = V^{-1/2} X$ and M is the orthogonal projection operator onto $C(B)$, and hence M is symmeric and idempotent. We have $y - X\hat{\beta} = (I - A)y = (I - V^{1/2} M V^{-1/2})y$.

Thus,

$$
\begin{aligned}
\Sigma_\gamma &= \text{Cov}(\hat{\gamma}) \\
&= \text{Cov}\left\{ D Z^T V^{-1} (Y - X\hat{\beta}) \right\} \\
&= \text{Cov}\left\{ D Z^T V^{-1} (I - V^{1/2} M V^{-1/2}) Y \right\} \\
&= D Z^T V^{-1} (I - V^{1/2} M V^{-1/2}) V (I - V^{-1/2} M V^{1/2}) V^{-1} Z D \\
&= D Z^T V^{-1} (I - V^{1/2} M V^{-1/2}) V^{1/2} V^{1/2} (I - V^{-1/2} M V^{1/2}) V^{-1} Z D \\
&= D Z^T V^{-1} (V^{1/2} - V^{1/2} M)(V^{1/2} - M V^{1/2}) V^{-1} Z D \\
&= D Z^T V^{-1/2} (I - M)(I - M) V^{-1/2} Z D \\
&= D Z^T V^{-1/2} (I - M) V^{-1/2} Z D \\
&= D Z^T V^{-1/2} (I - V^{-1/2} X (X^T V^{-1} X)^{-1} X^T V^{-1/2}) V^{-1/2} Z D \\
&= D Z^T (V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}) Z D
\end{aligned}
$$

where the fourth to last equality follows because M is an orthogonal projection operator, so $I - M$ is one, and hence is idempotent.

Thus, $\hat{\gamma} \sim N(0, \Sigma_\gamma)$.

2.b.3

Show that $\hat{\gamma}$ is the best linear unbiased predictor of γ.

Solution: The best linear unbiased predictor is defined as $\mathbb{E}(\gamma|Y)$. Note that

$$\begin{aligned}
\mathrm{Cov}(Y,\gamma) &= \mathrm{Cov}(X\beta + Z\gamma + \epsilon, \gamma) \\
&= \mathrm{Cov}(Z\gamma + \epsilon, \gamma) \\
&= \mathrm{Cov}(Z\gamma, \gamma) + \mathrm{Cov}(\epsilon, \gamma) \\
&= Z\mathrm{Cov}(\gamma)Z^T + 0 \\
&= ZDZ^T
\end{aligned}$$

Since $Y|\gamma$, and γ are multivariate normal, so is (Y,γ) and we have

$$\mathbb{E}Y|\gamma = 0 + ZDZ^T V^{-1}(Y - X\beta)$$

Substituting $\hat{\beta}$ in for β, we get that the BLUP is $ZDZ^T V^{-1}(Y - X\beta)$, which we identify as $\hat{\gamma}$.

2.c

Now suppose that R is of the form $R = \sigma^2 I_n$ where I_n is the $n \times n$ identity matrix, and (β, σ^2, D) are all unknown. Devise a detailed EM algorithm for jointly estimating (β, σ^2, D)

Solution:
The full data likelihood is

$$\begin{aligned}
&L(\beta, D, \sigma^2) \\
&\propto \det\{D\}^{-1/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta - Z\gamma)^T(Y - X\beta - Z\gamma) - \frac{1}{2}\gamma^T D^{-1}\gamma\right\} \\
&= \det\{D\}^{-1/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta - Z\gamma)^T(Y - X\beta - Z\gamma) - \frac{1}{2}\operatorname{tr}\big(D^{-1}\gamma\gamma^T\big)\right\}
\end{aligned}$$

The full data log likelihood is

$$\ell(\beta, D, \sigma^2) = -\frac{1}{2}\log\det\{D\} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta - Z\gamma)^T(Y - X\beta - Z\gamma) - \frac{1}{2}\operatorname{tr}\big(D^{-1}\gamma\gamma^T\big)$$

<u>M-step:</u>

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{2\sigma^2}X^T(Y - X\beta - Z\gamma) \overset{\text{SET}}{=} 0$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta - Z\gamma)^T(Y - X\beta - Z\gamma) \overset{\text{SET}}{=} 0$$

$$\begin{aligned}
\frac{\partial \ell}{\partial D} &= \frac{\partial \ell}{\partial D^{-1}}\frac{\partial D^{-1}}{\partial D} \\
&= \left(\frac{1}{2}D - \frac{1}{2}\gamma\gamma^T\right)\big(-D^{-1} \otimes D^{-1})\big) \overset{\text{SET}}{=} 0 \\
&\iff D - \gamma\gamma^T = 0
\end{aligned}$$

<u>E-step:</u>
We know that if D and σ^2 are known, $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ solves the likelihood equations by a previous part, where $V = Z^T D Z + \sigma^2 I$ Let $V_k = Z^T D_k Z + \sigma_k^2 I$. Then

$$\begin{aligned}
\mathbb{E}\frac{\partial \ell}{\partial \beta}|Y, \theta_k &= \frac{1}{2\sigma_k^2}X^T(Y - X\beta - Z\gamma_k) = 0 \\
&\implies \hat{\beta}_{k+1} = (X^T V_k^{-1} X)^{-1} X^T V_k^{-1} Y
\end{aligned}$$

We can obtain $\hat{\gamma}$ from the EBLUP, which we know solves the score equations by a previous part.

$$\hat{\gamma}_{k+1} = D_k^T V_k^{-1}(Y - X\hat{\beta}_k)$$

From the score equations, we can obtain D from

$$D_{k+1} = \gamma_k \gamma_k^T$$

And we have

$$\begin{aligned} \mathbb{E}\frac{\partial \ell}{\partial \sigma^2}|\beta_k, \gamma_k, D_k, Y &= -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}(Y - X\beta_k - Z\gamma_k)^T(Y - X\beta_k Z\gamma_k) = 0 \\ \implies \hat{\sigma}_{k+1}^2 &= \frac{1}{n}(Y - X\hat{\beta}_k - Z\hat{\gamma}_k)^T(Y - X\hat{\beta}_k - Z\hat{\gamma}_k) \end{aligned}$$

The full EM algorithm can be written as:

1. Begin with initial guess $(\hat{\beta}_1, \sigma_1^2, D_1)$
2. Compute parameter updates:
   (a) $\hat{\beta}_{k+1} = (X^T V_k^{-1} X)^{-1} X^T V_k^{-1} Y$
   (b) $\hat{\gamma}_{k+1} = D_k^T V_k^{-1}(Y - X\hat{\beta}_k)$
   (c) $\hat{D}_{k+1} = \hat{\gamma}_k \hat{\gamma}_k^T$
   (d) $\hat{V}_{k+1} = \hat{\sigma}_k^2 I + \hat{D}_k$
   (e) $\hat{\sigma}_{k+1}^2 = \frac{1}{n}(Y - X\hat{\beta}_k - Z\hat{\gamma}_k)^T(Y - X\hat{\beta}_k - Z\hat{\gamma}_k)$
3. Repeat step 2 until "convergence", obtaining an update from each previous iteration.

2.d

Next, consider the case that D, R, and β are unknown and that R has a general structure. Define $A = I_n - M$, where M is the orthogonal projection operator on the column space of X, and write $W = B'Y$ where $A = BB'$ and $B'B = I_n$. Consider estimation of the unknown parameters using the marginal distribution of $Y|X$ in (a).

2.d.1

Let $\hat{\beta}$ denote the MLE of β when (D, R) are fixed. Show that $\mathrm{Cov}(W, \hat{\beta}) = 0$.

Solution:
It can easily be shown that $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$, where $V = R + Z'DZ$. Now, we have

$$
\begin{aligned}
\mathrm{Cov}(W, \hat{\beta}) = 0 &\iff \mathrm{Cov}(B'Y, (X'V^{-1}X)^{-1}X'V^{-1}Y) = 0 \\
&\iff B'\mathrm{Cov}(Y)V^{-1}X(X'V^{-1}X)^{-1} = 0 \\
&\iff B'VV^{-1}X(X'V^{-1}X)^{-1} = 0 \\
&\iff B'X(X'V^{-1}X)^{-1} = 0 \\
&\iff C(B) \perp C(X(X'V^{-1}X)^{-1}) \\
&\iff C(BB') \perp C(X(X'V^{-1}X)^{-1}) \qquad\qquad \text{(since } C(B) = C(BB')\text{)} \\
&\iff BB'X(X'V^{-1}X)^{-1} = 0 \\
&\iff (I - M)X(X'V^{-1}X)^{-1} = 0 \\
&\iff 0 = 0
\end{aligned}
$$

where the last implication follows because $I - M$ is the orthogonal projection operator onto $C(X)^{\perp}$, so $(I - M)X = 0$. Hence, the result follows.

2.d.2

Note that

$$\mathbb{E}W = \mathbb{E}B'Y = B'\mathbb{E}Y = B'X\beta$$

$$\text{Cov}(W) = \text{Cov}(B'Y) = B'\text{Cov}(Y)B = B'VB$$

Since W is a linear form of a normal random variable, we have

$$f_W(w) = (2\pi)^{-1/2}\det\{B'VB\}^{-1/2}\exp\left\{(w - B'X\beta)^T(B'VB)^{-1}(w - B'X\beta)\right\}$$

For $w = B'y$, we have

$$\begin{aligned} f_W(B'y) &= (2\pi)^{-1/2}\det\{B'VB\}^{-1/2}\exp\left\{(B'(y - X\beta))^T(B'VB)^{-1}(B'(y - X\beta))\right\} \\ &= (2\pi)^{-1/2}\det\{B'VB\}^{-1/2}\exp\left\{(y - X\beta)^T B(B'VB)^{-1}B'(y - X\beta))\right\} \end{aligned}$$

2.d.3

Since the two variables are uncorrelated, we can estimate (R, D), from V in the likelihood of W. This is uncorrelated, so we can obtain the MLE and then substitute them back into the likelihood for β.

2.e