2010 Qualifying Exam Section 2

February 21, 2019

1 Problem 1

1.a

We have

$$\begin{aligned}
\text{logit}\{P(Y_i = 1|X_i)\} &= \log(P(X_i|Y_i = 1)P(Y_i = 1)) - \log(P(X_i|Y_i = 0)P(Y_i = 0)) \\
&= \log \pi_1 + \log\left(\exp\{-\frac{1}{2\sigma^2}(x_1 - \mu_1)^2\}\right) - \log(1 - \pi_1) - \log\left(\exp\{-\frac{1}{2\sigma^2}(x_i - \mu_0)^2\}\right) \\
&= \log \frac{\pi_1}{1 - \pi_1} - \frac{1}{2\sigma^2}(x_i^2 - 2\mu_1 x_i + \mu_1^2) + \frac{1}{2\sigma^2}(x_i^2 - 2\mu_0 x_i + \mu_0^2) \\
&= \log \frac{\pi_1}{1 - \pi_1} + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \frac{\mu_1 - \mu_0}{\sigma^2} x_i \\
&= \alpha_0 + \alpha_1 x_i
\end{aligned}$$

where $\alpha_0 = \log \frac{\pi_1}{1-\pi_1} + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}$ and $\alpha_1 = \frac{\mu_1 - \mu_0}{\sigma^2}$

1.b

Based on the model in part (a), we have $y_i|x_i \sim \text{Ber}(p_i)$, where

$$\text{logit}(p_i) = z_i^T \alpha \iff p_i = \frac{e^{z_i^T \alpha}}{1 + e^{z_i^T \alpha}}$$

Note that

$$\begin{aligned}
\frac{\partial p_i}{\partial \alpha} &= \frac{e^{z_i^T \alpha}(1 + e^{z_i^T \alpha}) z_i - (e^{z_i^T \alpha})^2 z_i}{(1 + e^{z_i^T \alpha})^2} \\
&= \frac{e^{z_i^T \alpha}}{1 + e^{z_i^T \alpha}} \frac{1}{1 + e^{z_i^T \alpha}} z_i \\
&= p_i(1 - p_i) z_i
\end{aligned}$$

Let ℓ_i be the i-th contribution to the log likelihood of $Y_i|X_i$. Then

$$\ell_i = y_i \log p_i + (1-y_i)\log(1-p_i)$$

We have

$$\begin{aligned} \frac{\partial \ell_i}{\partial p_i} &= \frac{y_i}{p_i} - \frac{1-y_i}{1-p_i} \\ &= \frac{y_i - p_i}{p_i(1-p_i)} \end{aligned}$$

Thus,

$$\frac{\partial \ell_i}{\partial \alpha} = \frac{\partial \ell_i}{\partial p_i}\frac{\partial p_i}{\partial \alpha} = \frac{y_i - p_i}{p_i(1-p_i)} p_i(1-p_i) z_i = (y_i - p_i) z_i$$

We have

$$\frac{\partial^2 \ell_i}{\partial \alpha \partial \alpha^T} = \frac{\partial^2 \ell_i}{\partial p_i \partial \alpha}\frac{\partial p_i}{\partial \alpha^T} - p_i(1-p_i) z_i z_i^T$$

Let ℓ represent the log likelihood based on all n observations. Then

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \alpha} = \sum_{i=1}^{n} (y_i - p_i) z_i = Z^T \boldsymbol{e}(\alpha)$$

where $\boldsymbol{e}(\alpha) = (y_1 - p_1(\alpha), \ldots, y_n - p_n(\alpha))^T$

Moreover,

$$-\frac{\partial^2 \ell}{\partial \alpha \partial \alpha^T} = -\sum_{i=1}^{n} \frac{\partial^2 \ell_i}{\partial \alpha \partial \alpha^T} = \sum_{i=1}^{n} p_i(1-p_i) z_i z_i^T = Z^T V(\alpha) Z$$

where $V(\alpha) = \mathrm{diag}\{p_i(\alpha)(1-p_i(\alpha))\}_{i=1}^{n}$

Thus, the Newton-Raphson algorithm is

$$\alpha^{(k+1)} = \alpha^{(k)} - [Z^T V(\alpha) Z]^{-1} Z^T \boldsymbol{e}(\alpha)\big|_{\alpha=\alpha^{(k)}}$$

1.c

We have

$$p(x_i,1) = p(x_i|y_i=1)p(y_i=1) = \pi_1(2\pi\sigma^2)^{1/2}\exp\left\{\frac{-1}{2\sigma^2}(x_i-\mu_1)^2\right\}$$
$$p(x_i,0) = p(x_i|y_i=0)p(y_i=0) = (1-\pi_1)(2\pi\sigma^2)^{1/2}\exp\left\{\frac{-1}{2\sigma^2}(x_i-\mu_0)^2\right\}$$

Thus, since $y_i \in \{0,1\}$ we can write the joint pdf as

$$p(x,y) = (2\pi\sigma^2)^{n/2}\pi_1^{\sum_{i=1}^n y_i}(1-\pi_1)^{\sum_{i=1}^n(1-y_i)}\exp\left\{-\frac{1}{2\sigma^2}\left[y_i(x_i-\mu_1)^2+(1-y_i)(x_i-\mu_0)^2\right]\right\}$$

Let $\gamma = \sigma^2$ Then the log-likelihood function can be written as

$$\ell(\theta) = \frac{n}{2}\log\gamma + \left(\sum_{i=1}^n y_i\right)\log\pi_1 + \left(n-\sum_{i=1}^n y_i\right)\log(1-\pi_1) - \frac{1}{2\gamma}\left[\sum_{i=1}^n y_i(x_i-\mu_1)^2 + \sum_{i=1}^n(1-y_i)(x-\mu_0)^2\right]$$

To find the MLE's we take the partial derivatives of the log likelihood with respect to each parameter.

$$\frac{\partial\ell}{\partial\pi_1} = \frac{\sum_{i=1}^n y_i}{\pi_1} - \frac{n-\sum_{i=1}^n y_i}{1-\pi_1} \overset{\text{SET}}{=} 0 \implies \sum_{i=1}^n y_i - n\pi_1 = 0 \implies \hat{\pi_1} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\frac{\partial\ell}{\partial\mu_0} = \frac{1}{\gamma}\sum_{i=1}^n(1-y_i)(x_i-\mu_0) \overset{\text{SET}}{=} 0 \implies \hat{\mu}_0 = \frac{\sum_{i=1}^n(1-y_i)x_i}{\sum_{i=1}^n(1-y_i)}$$

$$\frac{\partial\ell}{\partial\mu_1} = \frac{1}{\gamma}\sum_{i=1}^n y_i(x_i-\mu_1) \overset{\text{SET}}{=} 0 \implies \sum_{i=1}^n y_ix_i = \mu_1\sum_{i=1}^n y_i \implies \hat{\mu}_1 = \frac{\sum_{i=1}^n y_ix_i}{\sum_{i=1}^n y_i}$$

$$\frac{\partial\ell}{\partial\gamma} = -\frac{n}{2\gamma} + \frac{1}{2\gamma^2}\sum_{i=1}^n\left[y_i(x_i-\mu_1)^2+(1-y_i)(x-\mu_0)^2\right] \overset{\text{SET}}{=} 0$$
$$\implies \frac{n}{\gamma} = \frac{1}{\gamma^2}\sum_{i=1}^n\left[y_i(x_i-\mu_1)^2+(1-y_i)(x-\mu_0)^2\right]$$
$$\implies \hat{\gamma} = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n\left[y_i(x_i-\hat{\mu}_1)^2+(1-y_i)(x-\hat{\mu_0})^2\right]$$

To find the asymptotic covariance matrix, we take the inverse of the Fisher information matrix. This is done by finding the expectations of the negative of all of the second-order partial derivatives.

$$
\begin{aligned}
-\mathbb{E}\frac{\partial^2\ell}{\partial\pi_1^2} &= -\mathbb{E}\left\{-\frac{\sum_{i=1}^n y_i}{\pi_1^2} - \frac{n-\sum_{i=1}^n y_i}{1-\pi_1^2}\right\} \\
&= \frac{n\pi_1}{\pi_1^2} + \frac{n(1-\pi_1)}{(1-\pi_1)^2} \\
&= \frac{n}{\pi_1} + \frac{n}{1-\pi_1} \\
&= \frac{n}{\pi_1(1-\pi_1)}
\end{aligned}
$$

$$
-E\frac{\partial^2\ell}{\partial\mu_0^2} = -\mathbb{E}\left\{-\frac{1}{\gamma}\sum_{i=1}^n(1-y_i)\right\} = \frac{n(1-\pi_1)}{\gamma} = \frac{n(1-\pi_1)}{\sigma^2}
$$

Analagously, $-\mathbb{E}\frac{\partial^2\ell}{\partial\mu_1^2} = \dfrac{n\pi_1}{\sigma^2}$

$$
-\mathbb{E}\frac{\partial^2\ell}{\partial\gamma^2} = -\mathbb{E}\left\{\frac{n}{2\gamma^2} - \frac{1}{\gamma^3}\sum_{i=1}^n\left[y_i(x_i-\mu_1)^2 + (1-y_i)(x_i-\mu_0)^2\right]\right\}
$$

Note that

$$
\begin{aligned}
\mathbb{E}[y_1(x_1-\mu_1) + (1-y_1)(x_0-\mu_0)] &= \sum_{y=0}^{1}\int_{\mathcal{X}}\left\{y(x-\mu_1)^2 + (1-y)(x-\mu_0)^2\right\}p(x,y)dx \\
&= \int_{\mathcal{X}}(x-\mu_1)^2 p(x,1)dx + \int_{\mathcal{X}}(x-\mu_0)p(x,0)dx \\
&= \pi_1\sigma^2 + (1-\pi_1)\sigma^2 \\
&= \sigma^2 = \gamma
\end{aligned}
$$

Hence, since the (x_i, y_i) are i.i.d., we have

$$
-\mathbb{E}\frac{\partial^2\ell}{\partial\gamma^2} = -\frac{n}{2\gamma^2} + \frac{n}{\gamma^2} = \frac{n}{2\gamma^2} = \frac{n}{2\sigma^4}
$$

Looking at the mixed partials, note that $\frac{\partial\ell}{\partial\pi_1}$ only depends on π_1 so all mixed partials involving π_1 are 0. $\frac{\partial\ell}{\partial\mu_0}$ does not depend on μ_1 and likewise for $\frac{\partial\ell}{\partial\mu_1}$, so these mixed partials are also 0. Finally, note that

$$
\begin{aligned}
\mathbb{E}(1-y_i)(x_i-\mu_0) &= \int_{\mathcal{X}}\sum_{y=0}^{1}(1-y)(x_i-\mu_0)p(x,y)dx \\
&= \int_{\mathcal{X}}(x_i-\mu_0)p(x,0)dx \\
&= (1-\pi_1)\int_{\mathcal{X}}(x_i-\mu_0)p(x|y=0)dx \\
&= 0
\end{aligned}
$$

Since $p(x|y=0)$ is the density of a $N(\mu_0, \sigma^2)$ random variable. Thus, the expectation of the mixed partial of γ and μ_0 is 0, and analogously, so is that of γ and μ_1.

Thus, the Fisher information is diagonal

$$I_n(\theta) = n \text{ diag}\left\{\frac{1}{\pi_1(1-\pi_1)}, \frac{1-\pi_1}{\sigma^2}, \frac{\pi_1}{\sigma^2}, \frac{1}{2\sigma^4}\right\}$$

Thus, the asymptotic covariance matrix is

$$\Sigma \equiv \lim_{n\to\infty} [(1/n)\ I_n(\theta)]^{-1} = \text{diag}\left\{\pi_1(1-\pi_1), \frac{\sigma^2}{1-\pi_1}, \frac{\sigma^2}{\pi_1}, 2\sigma^4\right\}$$

And we have that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{d}} N_4(0, \Sigma)$$

1.d

$g : \mathbb{R}^4 \to \mathbb{R}^2$ and is given by

$$(\alpha_1, \alpha_2)^T = g(\theta) = g(\pi_1, \mu_0, \mu_1, \sigma^2) = \left(\log \frac{\pi_1}{1-\pi_1} + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \frac{\mu_1 - \mu_0}{\sigma^2} \right)^T$$

$$\begin{aligned}
\frac{\partial \alpha_1}{\partial \pi_1} &= \frac{1}{\pi_1} + \frac{1}{1-\pi_1} = \frac{1}{\pi_1(1-\pi_1)} \\
\frac{\partial \alpha_1}{\partial \mu_0} &= \frac{\mu_0}{\sigma^2} \\
\frac{\partial \alpha_1}{\partial \mu_1} &= -\frac{\mu_1}{\sigma^2} \\
\frac{\partial \alpha_1}{\partial \gamma} &= \frac{\mu_1^2 - \mu_0^2}{2\gamma^2} = \frac{\mu_1^2 - \mu_0^2}{2\sigma^4}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \alpha_2}{\partial \pi_1} &= 0 \\
\frac{\partial \alpha_2}{\partial \mu_0} &= -\frac{1}{\sigma^2} \\
\frac{\partial \alpha_2}{\partial \mu_1} &= \frac{1}{\sigma^2} \\
\frac{\partial \alpha_2}{\partial \gamma} &= \frac{\mu_0 - \mu_1}{\gamma^2} = \frac{\mu_0 - \mu_1}{\sigma^4}
\end{aligned}$$

Let $L = \begin{pmatrix} \frac{1}{\pi_1(1-\pi_1)} & \frac{\mu_0}{\sigma^2} & -\frac{\mu_1}{\sigma^2} & \frac{\mu_1^2-\mu_0^2}{2\sigma^4} \\ 0 & -\frac{1}{\sigma^2} & \frac{1}{\sigma^2} & \frac{\mu_0-\mu_1}{\sigma^4} \end{pmatrix}$. By the Delta Method,

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{\text{d}} N(0, L\Sigma L^T)$$

Let l_i denote the ith column of the L matrix. We have since Σ is diagonal,

$$\begin{aligned}
L\Sigma L^T &= \sum_{i=1}^{4} \Sigma_{ii} l_i l_i^T \\
&= \pi_1(1-\pi_1) \begin{pmatrix} \frac{1}{\pi_1^2(1-\pi_1)^2} & 0 \\ \cdot & 0 \end{pmatrix} + \frac{\sigma^2}{1-\pi_1} \begin{pmatrix} \frac{\mu_0^2}{\sigma^4} & -\frac{\mu_0}{\sigma^4} \\ \cdot & \frac{1}{\sigma^4} \end{pmatrix} \\
&+ \frac{\sigma^4}{\pi_1} \begin{pmatrix} \frac{\mu_1^2}{\sigma^4} & -\frac{\mu_1}{\sigma^4} \\ \cdot & \frac{1}{\sigma^4} \end{pmatrix} + 2\sigma^4 \begin{pmatrix} \frac{\mu_1^2-\mu_0^2}{4\sigma^8} & \frac{(\mu_1-\mu_0)^2(\mu_0-\mu_1)}{2\sigma^8} \\ \cdot & \frac{(\mu_1-\mu_0)^2}{\sigma^8} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\pi_1(1-\pi_1)} + \frac{\mu_0^2}{(1-\pi_1)\sigma^2} + \frac{\mu_1^2}{\pi_1\sigma^2} + \frac{(\mu_1^2-\mu_0^2)^2}{2\sigma^4} & -\frac{\mu_0}{(1-\pi_1)\sigma^2} - \frac{\mu_1}{\pi_1\sigma^2} + \frac{(\mu_1^2-\mu_0^2)(\mu_0-\mu_1)}{\sigma^4} \\ \cdot & \frac{1}{(1-\pi_1)\sigma^2} + \frac{1}{\pi_1\sigma^2} + \frac{2(\mu_1-\mu_0)^2}{\sigma^4} \end{pmatrix}
\end{aligned}$$

2 Problem 2

2.a

Let $\beta = (\beta_1, \ldots, \beta_p)^T$ and let $X_i = (X_{i1}, \ldots, X_{ij})^T$. Then we can write

$$Y_i = X_i^T \beta + U + \epsilon_i$$

As a result, we have

$$\mathbb{E}Y_i = \mathbb{E}\left\{X_i^T \beta + U + \epsilon_i\right\} = X_i^T \beta + \mathbb{E}U + \mathbb{E}\epsilon_i = X_i^T \beta + \alpha$$

Moreover,

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_j) &= \mathrm{Cov}(X_i^T \beta + U + \epsilon_i, X_j^T \beta + U + \epsilon_i) \\
&= \mathrm{Cov}(U + \epsilon_i, U + \epsilon_j) \\
&= \mathrm{Cov}(U, U) + \mathrm{Cov}(\epsilon_i, U) + \mathrm{Cov}(U, \epsilon_j) + \mathrm{Cov}(\epsilon_i, \epsilon_j) \\
&= k\sigma^2 + \sigma^2 \mathbf{1}_{\{i=j\}}
\end{aligned}$$

Let Σ be the covariance matrix of Y and let Σ_{ij} denote its ij-th element. Then

$$\Sigma_{ij} = \begin{cases} (k+1)\sigma^2 & , i = j \\ k\sigma^2 & , i \neq j \end{cases}$$

It follows that we can write $\Sigma = \sigma^2 I + k\sigma^2 J = \sigma^2(I + kJ)$ By the hint, we have that Σ is nonsingular. Σ is also positive semidefinite since all covariance matrices are positive semidefinite. Since Σ is positive semidefinite and nonsingular, Σ is positive definite. A formal argument is below:

By the hint, Σ is invertible. Thus, its determinant is nonzero. Since the determinant is equal to the product of the eigenvalues, all of the eigenvalues of Σ are nonzero. Since Σ is a covariance matrix, it is positive semidefinite. Thus, all the eigenvalues of Σ are non-negative. But since we have just argued the eigenvalues of Σ are nonzero, all the eigenvalues of Σ must be positive. Thus, Σ is positive definite.

2.b

Note we can write the model as

$$Y_i = \alpha + X_{i1}\beta_1 + \ldots + X_{ip}\beta_p + U^* + \epsilon_i$$

where $U^* \sim N(0, k\sigma^2)$. Converting back to matrix notation, let $X^* = (\mathbf{1}, X_1, \ldots, X_p)$ be the design matrix and let $\beta^* = (\alpha, \beta)^T$. We can express the model as

$$Y = X^*\beta^* + \delta$$

where $\delta \sim N(0, \Sigma)$, $\Sigma = \sigma^2(I + kJ)$

As stated in the problem, the $n \times p$ matrix consisting of the elements of $X_{ij} - \bar{X}_i$ is full rank (has rank p). Let Z be this matrix and let z_i be the i-th column of Z. Let $\boldsymbol{x}_i$ denote the i-th column of the original design matrix, that is, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$. Then

$$z_i = \boldsymbol{x}_i - \bar{x}_i\mathbf{1} \quad i = 1, \ldots, n$$

Note that

$$\mathbf{1}^T z_i = \sum_{j=1}^{p} z_{ij} = \sum_{j=1}^{p}(x_{ij} - \bar{x}_{.j}) = 0$$

Hence, $\mathbf{1} \perp z_j$ for every j, so $\mathbf{1} \perp C(Z)$. Thus, $(1, z_1, \ldots, z_p)$ form a set of $(p+1)$ linearly independent vectors since Z is an $n \times p$ matrix of rank p and orthogonality implies linear independence. It follows that $(\mathbf{1}, Z)$ is a full rank $n \times (p+1)$ matrix.

We must show that $(\mathbf{1}, X)$ is a full rank matrix. Suppose $w \in C(\mathbf{1}, Z)$. Then $w = (1, Z)b$ for some $b = (b_0, b_1, \ldots, b_p)^T$. Thus,

$$\begin{aligned}
w &= (\mathbf{1}, Z)b \\
&= b_0\mathbf{1} + \sum_{j=1}^{p} b_j\boldsymbol{z}_j \\
&= b_0\mathbf{1} + \sum_{j=1}^{p} b_j(\boldsymbol{x}_j - \bar{x}_{.j}\mathbf{1}) \\
&= b_0\mathbf{1} + \sum_{j=1}^{p} b_j\boldsymbol{x}_j - \sum_{j=1}^{p} b_j\bar{x}_{.j}\mathbf{1} \\
&= (b_0 - \sum_{j=1}^{p} b_j\bar{x}_{.j})\mathbf{1} + \sum_{j=1}^{p} b_j\boldsymbol{x}_j \\
&\in C(1, X)
\end{aligned}$$

and thus, $C(1, Z) \subset C(1, X) \implies r(1, X) \geq r(1, Z) = p + 1$ So we have that $(1, X)$ is a $n \times (p+1)$ matrix and thus $r(1, X) \leq p + 1$. Thus, we must have $r(1, X) = p + 1$ and hence $(1, X)$ is full rank.

Since X^* is full rank, β^* is estimable because the columns of X^* form a basis for R^{p+1}, and hence we can find a P^T such that $I_p = P^T X^*$

2.c

Let $W = (I + kJ)$ so that $\Sigma = \sigma^2 W$ and to ease notation let $Z = X^*$ denote the design matrix. Then

$$L_n(\theta, \sigma^2) = \det\left\{\sigma^2 W\right\} \exp\left\{-\frac{1}{2\sigma^2}(y - Z\theta)^T W^{-1}(y - Z\theta)\right\}$$

$$\begin{aligned}\ell_n(\theta, \sigma^2) &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - Z\theta)^T W^{-1}(y - Z\theta) \\ &= -\frac{n}{2}\log\sigma^2 - \frac{1}{\sigma^2}\left(\frac{1}{2}y^T W^{-1} y + \theta^T Z^T W^{-1} y - \frac{1}{2}\theta^T Z^T W^{-1} Z\theta\right)\end{aligned}$$

$$\frac{\partial \ell_n}{\partial \theta} = \frac{1}{\sigma^2}(Z^T W^{-1} y - Z^T W^{-1} Z\theta) \overset{\text{SET}}{=} 0$$

$$\implies \hat{\theta} = (Z^T W^{-1} Z)^{-1} Z^T W^{-1} y$$

Note that

$$\mathbb{E}\hat{\theta} = \mathbb{E}(Z^T W^{-1} Z)^{-1} Z^T W^{-1} y = (Z^T W^{-1} Z)^{-1} Z^T W^{-1}\mathbb{E}y = (Z^T W^{-1} Z)^{-1} Z^T W^{-1} Z\theta = \theta$$

$$\begin{aligned}\operatorname{Cov}(\hat{\theta}) &= \operatorname{Cov}((Z^T W^{-1} Z)^{-1} Z^T W^{-1} y) \\ &= (Z^T W^{-1} Z)^{-1} Z^T W^{-1} \operatorname{Cov}(y) W^{-1} Z (Z^T W^{-1} Z)^{-1} \\ &= (Z^T W^{-1} Z)^{-1} Z^T W^{-1} (\sigma^2 W) W^{-1} Z (Z^T W^{-1} Z)^{-1} \\ &= \sigma^2 (Z^T W^{-1} Z)^{-1} Z^T W^{-1} W W^{-1} Z (Z^T W^{-1} Z)^{-1} \\ &= \sigma^2 (Z^T W^{-1} Z)^{-1}\end{aligned}$$

Since $\hat{\theta}$ is a linear function of y, which is multivariate normal, we have

$$\hat{\theta} \sim N_{p+1}(\theta, \sigma^2 (Z^T W^{-1} Z)^{-1})$$

2.d

We can write the sum of squares as

$$Q = (y - Z\theta)^T(y - Z\theta) = y^Ty - 2\theta^TZ^Ty + \theta^TZ^TZ\theta$$

We have

$$\frac{\partial Q}{\partial \theta} = -2Z^Ty + 2Z^TZ\theta$$

which equals 0 if and only if $\tilde{\theta} = (Z^TZ)^{-1}Z^Ty$

Now, consider the transformation

$$y^* = W^{-1/2}y = W^{-1/2}Z\theta + W^{-1/2}\epsilon = Z^*\theta + \epsilon^*$$

where $\epsilon^* = W^{-1/2}\epsilon \sim N(0, W^{-1/2}\sigma^2WW^{-1/2}) =_d N(0, \sigma^2I)$.

By the Gauss Markov theorem, the BLUE is given by

$$\hat{\beta}^* = (Z^{*T}Z^*)^{-1}Z^{*T}y^* = (Z^TW^{-1}Z)^{-1}Z^TW^{-1}y = \hat{\beta}$$

Note that $\hat{\theta}$ is a linear estimator and by part (b) α is estimable, so there exists some ρ such that $\lambda' = \rho'X$ where $\lambda'\theta = \alpha$. By the Gauss Markov Theorem, we have

$$\text{Var}(\tilde{\alpha}) = \text{Var}(a'y^*) = \text{Var}(a'y^* - \rho'My^* + \rho'My^*)$$

Since $\hat{\theta}$ is unbiased for θ, we know that $\text{Cov}(a'y^* - \rho'My^*, \rho'My^*) = 0$. Thus, we have

$$\text{Var}(\tilde{\alpha}) = \text{Var}(\rho'My^*) + \text{Var}(a'y^* - \rho'My^*) = \text{Var}(\hat{\alpha}) + \text{Var}(a'y^* - \rho'My^*)$$

So if the variance is larger, we need to show $\text{Var}(a'y - \rho'My) > 0$. Note that

$$\begin{aligned}
\text{Var}(a'y^* - \rho'My^*) = 0 &\iff (a^T - \rho^TM)\text{Cov}(y^*)(a^T - \rho^TM)^T = 0 \\
&\iff (a - M\rho)^T\text{Cov}(y^*)(a - M\rho) = 0 \\
&\iff (a - M\rho)^T\sigma^2I(a - M\rho) = 0 \\
&\iff \sigma^2(a - M\rho)^T(a - M\rho) = 0 \\
&\iff ||a - M\rho||^2 = 0 \\
&\iff a - M\rho = 0 \\
&\iff a = M\rho
\end{aligned}$$

Now, since Z is full rank, $\tilde{\alpha} = \lambda^T\tilde{\theta}$ and $\hat{\alpha} = \lambda^T\hat{\theta}$. These two estimators are clearly different, so $a \neq M\rho$. Thus, we have $\text{Var}(\hat{\alpha}) < \text{Var}(\tilde{\alpha})$

3 Problem 3

3.a

Note that

$$\begin{aligned} AUC(\beta) &= P(\beta^T X_1 \geq \beta^T Y_1) \\ &= P(\beta^T (X_1 - Y_1) \geq 0) \end{aligned} \tag{3.1}$$

Now, $\mathbb{E}\beta^T(X_1 - Y_1) = \beta^T(\mathbb{E}X_1 - \mathbb{E}Y_1) = \beta^T(\mu_1 - \mu_2)$ and since X_1 and Y_1 are independent,

$\text{Var}(\beta^T(X_1 - Y_1)) = \text{Var}(\beta^T X_1 \beta - \beta^T Y_1 \beta) = \text{Var}(\beta^T X_1 \beta) + \text{Var}(\beta^T Y_1 \beta) = \beta^T \Sigma\beta + \beta^T \Sigma\beta = 2\beta^T \Sigma\beta$

Thus, since X_1 and Y_1 are independent bivariate normal and $\beta^T(X_1 - Y_1)$ is a linear function of them, we have

$$\beta^T(X_1 - Y_1) \sim N(\beta^T(\mu_1 - \mu_2), 2\beta^T\Sigma\beta)$$

By the relationship in (3.1), we have

$$\begin{aligned} AUC &= P\left(\frac{\beta^T(X_1 - Y_1) - \beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}} \geq -\frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) \\ &= P\left(N(0,1) \geq -\frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) \\ &= P\left(N(0,1) \leq \frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) \\ &= \Phi\left(\frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) \end{aligned}$$

where the third equality follows by the symmetry of the normal pdf about 0.

3.b

Let $Q(\beta) = \frac{1}{\sqrt{2}} \frac{\beta^T(\mu_1 - \mu_2)}{(\beta^T \Sigma \beta)^{1/2}}$. Since AUC is a a monotone increasing function of $Q(\beta)$, optimizing AUC is equivalent to optimizing $Q(\beta)$.

We have

$$Q'(\beta) = \frac{(\mu_1 - \mu_2)(\beta^T \Sigma \beta^{1/2}) - \beta^T(\mu_1 - \mu_2)(\beta^T \Sigma \beta)^{-1/2} \Sigma \beta}{\sqrt{2} \beta^T \Sigma \beta} \overset{\text{SET}}{=} 0$$

$$\iff \beta^T(\mu_1 - \mu_2)(\beta^T \Sigma \beta)^{-1/2} \Sigma \beta = (\beta^T \Sigma \beta)^{1/2}(\mu_1 - \mu_2)$$

$$\iff \frac{\beta^T(\mu_1 - \mu_2) \Sigma \beta}{\beta^T \Sigma \beta} = \mu_1 - \mu_2$$

Let $\hat{\beta} = \Sigma^{-1}(\mu_1 - \mu_2)$. Then

$$LHS = \frac{(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) \Sigma \Sigma^{-1}(\mu_1 - \mu_2)}{(\mu_1 - \mu_2)^T \Sigma (\mu_1 - \mu_2)} = \mu_1 - \mu_2$$

And hence $\hat{\beta} = \Sigma^{-1}(\mu_1 - \mu_2)$ is optimal, corresponding to

$$Q(\hat{\beta}) = \frac{1}{\sqrt{2}}(\mu_1 - \mu_2)^T \Sigma (\mu_1 - \mu_2) / [(\mu_1 - \mu_2)^T \Sigma (\mu_1 - \mu_2)]^{1/2} = \left(\frac{(\mu_1 - \mu_2)^T \Sigma (\mu_1 - \mu_2)}{2} \right)^{1/2}$$

ALTERNATE SOLUTION: By the Cauchy-Schwartz Inequality,

$$\begin{aligned}
Q(\beta) &= \frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2 \beta^T \Sigma \beta}} \\
&= \frac{\beta^T \Sigma^{1/2} \Sigma^{-1/2}(\mu_1 - \mu_2)}{\sqrt{2 \beta^T \Sigma^{1/2} \Sigma^{1/2} \beta}} \\
&= \frac{(\Sigma^{1/2}\beta)^T(\Sigma^{-1/2}(\mu_1 - \mu_2))}{\sqrt{2}||\Sigma^{1/2}\beta||} \\
&\leq \frac{||\Sigma^{1/2}\beta|| \times ||\Sigma^{-1/2}(\mu_1 - \mu_2)||}{\sqrt{2}||\Sigma^{1/2}\beta||} \\
&= \sqrt{\frac{(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)}{2}}
\end{aligned}$$

Thus, $\beta = \Sigma^{-1}(\mu_1 - \mu_2)$ is the maximum since it attains the upper bound.

3.c

The likelihood equation is given by

$$
\begin{aligned}
L(\mu_1, \mu_2, \Sigma) &= \prod_{i=1}^{n} L(\mu_1, \Sigma | x_i) \times \prod_{j=1}^{m} L(\mu_2, \Sigma | y_j) \\
&= \prod_{i=1}^{n} \left\{ \det\{\Sigma\}^{-1/2} \exp\left\{ -\frac{1}{2}(x_i - \mu_1)^T \Sigma^{-1} (x_i - \mu_1) \right\} \right\} \\
&\quad \times \prod_{j=1}^{m} \left\{ \det\{\Sigma\}^{-1/2} \exp\left\{ -\frac{1}{2}(y_j - \mu_2)^T \Sigma^{-1} (y_j - \mu_2) \right\} \right\} \\
&= \det\{\Sigma\}^{-\frac{n+m}{2}} \exp\left\{ -\frac{1}{2} \left(\sum_{i=1}^{n} [(x_i - \mu_1)^T \Sigma^{-1} (x_i - \mu)] + \sum_{j=1}^{m} [y_i - \mu_2)^T \Sigma^{-1} (y_i - \mu_2)] \right) \right\}
\end{aligned}
$$

The log-likelihood is given by

$$
\begin{aligned}
\ell(\mu_1, \mu_2, \Sigma) &= -\frac{n+m}{2} \log\det\{\Sigma\} - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_1)^T \Sigma^{-1} (x_i - \mu_1) \\
&\quad - \frac{1}{2} \sum_{j=1}^{m} (y_j - \mu_2)^T \Sigma^{-1} (y_j - \mu_2) \qquad (3.2) \\
&= -\frac{n+m}{2} \log\det\{\Sigma\} - \frac{1}{2} \operatorname{tr}\left\{ \Sigma^{-1} \sum_{i=1}^{n} (x_i - \mu_1)(x_i - \mu_1)^T \right\} \\
&\quad - \frac{1}{2} \operatorname{tr}\left\{ \Sigma^{-1} \sum_{j=1}^{m} (y_j - \mu_2)(y_j - \mu_2)^T \right\} \qquad (3.3)
\end{aligned}
$$

$$
\frac{\partial \ell}{\partial \mu_1} = \Sigma^{-1} \sum_{i=1}^{n} (x_i - \mu_1)
$$

which is 0 if and only if $\hat{\mu}_1 = \bar{x}$. Similarly, we have $\hat{\mu}_2 = \bar{y}$.

Note that

$$
\frac{\partial}{\partial \Sigma_{ij}} \log\det\{\Sigma\} = \frac{1}{\det\{\Sigma\}} \frac{\partial \det\{\Sigma\}}{\partial \Sigma_{ij}} = \frac{1}{\det\{\Sigma\}} \operatorname{adj}(\Sigma)_{ji} = (\Sigma^{-1})_{ji} = (\Sigma^{-1})_{ij}
$$

and thus we have $\frac{\partial}{\partial \Sigma} \log\det\{\Sigma\} = \Sigma^{-1}$.

Note also that

$$
\frac{\partial \operatorname{tr}(AB)}{\partial A} = B^T
$$

So by relationship (3.3), we have

$$\frac{\partial \ell}{\partial \Sigma^{-1}} = \frac{n+m}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu_1)(x_i - \mu_2)^T - \frac{1}{2}\sum_{j=1}^{n}(y_i - \mu_2)(y_i - \mu_2)^T$$

Setting this equal to 0, we get

$$\hat{\Sigma} = \left(\frac{1}{n+m}\sum_{i=1}^{n}(x_i - \mu_1)(x_i - \mu_2)^T + \sum_{j=1}^{m}(y_j - \mu_2)^T(y_j - \mu_2)\right)$$

By the invariance property of MLEs, we have

$$\hat{A} = \Phi\left(\left[(\hat{\mu}_1 - \hat{\mu}_2)^T\hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2)/2\right]^{1/2}\right)$$

3.d

One can use the Delta method directly here or use the theory of maximum likelihood theory to get the asymptotic distribution, but you do not have to. Let $\hat{\eta} = \hat{\mu}_1 - \hat{\mu}_2$ and let $\eta = \mu_1 - \mu_2$. Focusing on what is inside the Φ, we have

$$\begin{aligned}
\sqrt{n}(\hat{\eta}'\hat{\Sigma}^{-1}\hat{\eta} - \eta'\Sigma\eta) &= \sqrt{n}\left\{(\hat{\eta} - \eta + \eta)'\hat{\Sigma}^{-1}(\hat{\eta} - \eta + \eta) - \eta'\Sigma^{-1}\eta\right\} \\
&= \sqrt{n}\left\{(\hat{\eta} - \eta)'\hat{\Sigma}^{-1}(\hat{\eta} - \eta) + 2\eta'\hat{\Sigma}^{-1}(\hat{\eta} - \eta) + \eta'\hat{\Sigma}^{-1}\eta - \eta^T\Sigma^{-1}\eta\right\} \\
&= \sqrt{n}(\hat{\eta} - \eta)'\hat{\Sigma}^{-1}(\hat{\eta} - \eta) + 2\sqrt{n}\eta'\hat{\Sigma}^{-1}(\hat{\eta} - \eta) + \sqrt{n}\eta'(\hat{\Sigma}^{-1} - \Sigma^{-1})\eta \\
&= A_n + B_n + C_n
\end{aligned}$$

Now, by the WLLN, $\hat{\Sigma}$ is consistent for Σ, and so by the continuous mapping theorem, $\hat{\Sigma}^{-1}$ is consistent for Σ^{-1}. Thus, we can write $\hat{\Sigma}^{-1} = \Sigma^{-1}(1 + o_p(1))$. We have

$$\begin{aligned}
A_n &= \sqrt{n}(\hat{\eta} - \eta)'\Sigma^{-1}(1 + o_p(1))(\hat{\eta} - \eta) \\
&= \sqrt{n}(\hat{\eta} - \eta)'\Sigma^{-1}(\hat{\eta} - \eta) + o_p(1)\sqrt{n}(\hat{\eta} - \eta)'(\hat{\eta} - \eta)
\end{aligned}$$

By the central limit theorem and the continuous mapping theorem, $\Sigma^{-1/2}\sqrt{n}(\hat{\eta} - \eta) \to N_2(0, (1+\tau^2)I)$ in distribution. Hence, $n(\hat{\eta} - \eta)'\Sigma^{-1}(\hat{\eta} - \eta) = O_p(1)$, so $(\hat{\eta} - \eta)'\Sigma^{-1}(\hat{\eta} - \eta) = O_p(1/n)$ and so the left hand side of A_n is $o_p(1)$. Similarly, we have the right-hand side of A_n is $o_p(1)$. Thus, $A_n = o_p(1)$.

Let $\mu = (\mu_1, \mu_2)$ and $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)$. By the multivariate version of the CLT and by the above,

$$\sqrt{n}(\hat{\mu} - \mu) = \sqrt{n}\begin{pmatrix} \hat{\mu_1} - \mu_1 \\ \tau(\hat{\mu_2} - \mu_2) \end{pmatrix} \xrightarrow{\text{d}} N_4\left(0, \begin{pmatrix} \Sigma & 0 \\ 0 & \tau^2\Sigma \end{pmatrix}\right)$$

where the covariance matrix is block diagonal because the x's and y's are independent.

Hence, by the continuous mapping theorem, we have

$$\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{\text{d}} N_2(0, (1 + \tau^2), \Sigma)$$

Thus, by Slutsky's theorem, we have

$$\begin{aligned}
B_n &= 2\sqrt{n}\eta'\hat{\Sigma}^{-1}(\hat{\eta} - \eta) \xrightarrow{\text{d}} 2\eta' N_2(0, (1 + \tau^2), \Sigma^{-1/2}) \\
&\overset{\text{d}}{=} N_1(0, 4(1 + \tau)^2\eta^T\Sigma^{-1/2}\eta)
\end{aligned}$$

Finally,

$$C_n = \sqrt{n}(\eta^T\widehat{\Sigma^{-1}}\eta - \eta^T\Sigma^{-1}\eta) \to N(0, \delta^2)$$

in distribution, where $\delta^2 = n\mathrm{Var}(\eta_1^2(\hat{\Sigma}^{-1})_{11} + \eta_2^2(\hat{\Sigma}^{-1})_{22} + 2\eta_{12}(\hat{\Sigma}^{-1})_{12})$

As a result, the whole thing converges to a normal distribution, and then we can apply the function Φ using the delta method to get the final result.