

UCLA Biostatistics 250C

Final Exam

Due Date: 11:59pm, June 19th, 2021

Time: Take-Home

Student: Elvis Cui

June 13, 2021

Problem 1

(a)

Note that we are only interested in θ :

$$\begin{aligned}
 p(\theta|\Sigma, \mathbf{x}) &\propto p(\mathbf{x}, \theta, \Sigma) \\
 &\propto p(\mathbf{x}, \theta|\Sigma) \\
 &= \mathcal{N}(\mathbf{x}|\theta, \Sigma) \times \mathcal{N}(\theta|\mu, \rho\Sigma) \\
 &\propto \exp\left(-\frac{1}{2}(\mathbf{x} - \theta)^T \Sigma^{-1}(\mathbf{x} - \theta)\right) \times \exp\left(-\frac{1}{2\rho}(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)\right) \\
 &= \exp\left(-\frac{1}{2}\left[\theta^T\left((1 + \frac{1}{\rho})\Sigma^{-1}\right)\theta - 2(\mathbf{x} + \frac{\mu}{\rho})^T \Sigma^{-1}\theta + \mathbf{c}\right]\right)
 \end{aligned}$$

where \mathbf{c} is a constant that does not depend on θ .

Next, apply Dr. Banerjee's Mm -formula,

$$\begin{aligned}
 \mathbf{M}^{-1} &= \left(1 + \frac{1}{\rho}\right) \Sigma^{-1} \\
 \mathbf{m} &= \Sigma^{-1} \left(\mathbf{x} + \frac{\mu}{\rho}\right)
 \end{aligned}$$

Hence,

$$\mathbf{M} = \frac{\rho}{\rho + 1} \Sigma \quad (1)$$

$$\mathbf{Mm} = \frac{\rho}{\rho + 1} \left(\mathbf{x} + \frac{\mu}{\rho}\right) \quad (2)$$

$$p(\theta|\Sigma, \mathbf{x}) =_{a.s.} \mathcal{N}(\theta|\mathbf{Mm}, \mathbf{M}) \quad (3)$$

(b)

By part (a), we have

$$\begin{aligned}
 \mathbf{c} &= \frac{1}{\rho} \mu^T \Sigma^{-1} \mu + \mathbf{x}^T \Sigma^{-1} \mathbf{x} - \mathbf{m}^T \mathbf{Mm} \\
 &= \text{Tr} \left(\left(\frac{1}{\rho} \mu \mu^T + \mathbf{x} \mathbf{x}^T - \frac{\rho}{\rho + 1} \left(\mathbf{x} + \frac{\mu}{\rho}\right) \left(\mathbf{x} + \frac{\mu}{\rho}\right)^T \right) \Sigma^{-1} \right) \\
 &= \text{Tr} \left(\frac{1}{\rho + 1} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \right) \Sigma^{-1}
 \end{aligned}$$

Next, by collapsing or marginalizing,

$$p(\Sigma|\mathbf{x}) = \int_{\Theta} \underbrace{\mathcal{N}(\mathbf{x}|\boldsymbol{\theta}, \Sigma)\mathcal{N}(\boldsymbol{\theta}|\Sigma)\mathcal{IW}(\Sigma)}_{(*)} d\boldsymbol{\theta}$$

But from part (a) and previous argument, we know that

$$(*) \propto |\Sigma|^{-\frac{\alpha+p+3}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \mathbf{M}\mathbf{m})^T \mathbf{M}^{-1}(\boldsymbol{\theta} - \mathbf{M}\mathbf{m}) - \frac{\mathbf{c}}{2} + \frac{1}{2}\text{Tr}(\boldsymbol{\Omega}\Sigma^{-1})\right)$$

Integrating out $\boldsymbol{\theta}$ will result in an additional factor that is proportional to $|\Sigma|^{\frac{1}{2}}$. Hence,

$$p(\Sigma|\mathbf{x}) \propto |\Sigma|^{-\frac{\alpha+1+p+1}{2}} \exp\left(-\frac{1}{2}\text{Tr}\left(\boldsymbol{\Omega} + \frac{1}{\rho+1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\right)\Sigma^{-1}\right) \quad (4)$$

Or equivalently,

$$p(\Sigma|\mathbf{x}) =_{a.s.} \mathcal{IW}\left(\Sigma \middle| \alpha + 1, \boldsymbol{\Omega} + \frac{1}{\rho+1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\right) \quad (5)$$

(c)

We can draw samples using **method of mixtures** or **composition sampling**:

- Collect observed data \mathbf{x} and hyper-parameters $(\alpha, \rho, \boldsymbol{\mu})$.
- Sample Σ from

$$\mathcal{IW}\left(\Sigma \middle| \alpha + 1, \frac{1}{\rho+1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\right)$$

which depends on Inverse Wishart generators. Call the generated samples

$$\tilde{\Sigma} = (\Sigma_1, \Sigma_2, \dots, \Sigma_n)$$

- Sample $\boldsymbol{\theta}$ from

$$\mathcal{N}(\mathbf{M}_i \mathbf{m}, \mathbf{M}_i)$$

where $\mathbf{M}_i = \rho/(\rho+1)\Sigma_i$ and the above depends only on multivariate Normal generators.

Call the generated samples

$$\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n)$$

- Output the generated samples

$$(\tilde{\Sigma}, \tilde{\boldsymbol{\theta}})$$

(d)

For any bounded measurable function $f(\boldsymbol{\theta})$, we have

$$\mathbb{E}[f(\boldsymbol{\theta})|\mathbf{x}] = \mathbb{E}[\mathbb{E}[f(\boldsymbol{\theta})|\boldsymbol{\Sigma}, \mathbf{x}]|\mathbf{x}]$$

The outer conditional expectation can be approximated by

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[f(\boldsymbol{\theta})|\boldsymbol{\Sigma}_i, \mathbf{x}] =_{a.s.} \mathbb{E}[f(\boldsymbol{\theta})|\mathbf{x}]$$

To approximate the inner conditional expectation $\mathbb{E}[f(\boldsymbol{\theta})|\boldsymbol{\Sigma}_i, \mathbf{x}]$, one uses method of moments:

$$\mathbb{E}[f(\boldsymbol{\theta})|\boldsymbol{\Sigma}_i, \mathbf{x}] \approx f(\boldsymbol{\theta}_i)$$

Or more correctly (if computational power permits),

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{j=1}^m f(\boldsymbol{\theta}_{ij}) =_{a.s.} \mathbb{E}[f(\boldsymbol{\theta})|\boldsymbol{\Sigma}_i, \mathbf{x}]$$

Hence, for a double array

$$\begin{aligned} &\{\boldsymbol{\theta}_{11}, \boldsymbol{\theta}_{12}, \dots, \boldsymbol{\theta}_{1m}, \\ &\quad \boldsymbol{\theta}_{21}, \boldsymbol{\theta}_{12}, \dots, \boldsymbol{\theta}_{2m}, \\ &\quad \dots\dots\dots, \\ &\quad \boldsymbol{\theta}_{n1}, \boldsymbol{\theta}_{n2}, \dots, \boldsymbol{\theta}_{nm}\} \end{aligned}$$

We have

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{m} \sum_{j=1}^m f(\boldsymbol{\theta}_{ij}) \right) \right) =_{a.s.} \mathbb{E}[f(\boldsymbol{\theta})|\mathbf{x}] \quad (6)$$

In other words, if we first draw $\boldsymbol{\Sigma}_i$ from $p(\boldsymbol{\Sigma}|\mathbf{x})$ and then draw $\boldsymbol{\theta}_i$ from $p(\boldsymbol{\theta}|\boldsymbol{\Sigma}_i, \mathbf{x})$, the resulting $\boldsymbol{\theta}_i$ has marginal distribution $p(\boldsymbol{\theta}|\mathbf{x})$.

Problem 2

(a)

By the definition of covariance, we have

$$\text{Cov}(\mathbf{y}) = \text{Cov} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \dots \\ \mathbf{y}_n \end{pmatrix} \quad (7)$$

$$= \begin{pmatrix} \text{Cov}(\mathbf{y}_1, \mathbf{y}_1) & \text{Cov}(\mathbf{y}_1, \mathbf{y}_2) & \dots & \text{Cov}(\mathbf{y}_1, \mathbf{y}_n) \\ \text{Cov}(\mathbf{y}_2, \mathbf{y}_1) & \text{Cov}(\mathbf{y}_2, \mathbf{y}_2) & \dots & \text{Cov}(\mathbf{y}_2, \mathbf{y}_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\mathbf{y}_n, \mathbf{y}_1) & \text{Cov}(\mathbf{y}_n, \mathbf{y}_2) & \dots & \text{Cov}(\mathbf{y}_n, \mathbf{y}_n) \end{pmatrix} \quad (8)$$

$$= \begin{pmatrix} a_{11}\mathbf{\Lambda} & a_{12}\mathbf{\Lambda} & \dots & a_{1n}\mathbf{\Lambda} \\ a_{21}\mathbf{\Lambda} & a_{22}\mathbf{\Lambda} & \dots & a_{2n}\mathbf{\Lambda} \\ \dots & \dots & \dots & \dots \\ a_{n1}\mathbf{\Lambda} & a_{n2}\mathbf{\Lambda} & \dots & a_{nn}\mathbf{\Lambda} \end{pmatrix} \quad (9)$$

$$= \mathbf{A} \otimes \mathbf{\Lambda} \quad (10)$$

where \otimes is the Kronecker product.

(b) and (c)

Here I quote 2 results on Kronecker product and $\text{Vec}(\cdot)$ operator. One is from lecture 17, May 26th and the other is from lecture 18, June 2nd.

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \\ (\mathbf{A} \otimes \mathbf{B}) \text{Vec}(\mathbf{X}) &= \text{Vec}(\mathbf{B}\mathbf{X}\mathbf{A}^T) \end{aligned}$$

Next, let

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$$

be the matrix version of the stacked vector \mathbf{y} . Then we have

$$\begin{aligned} \mathbf{y}^T (\mathbf{A} \otimes \mathbf{\Lambda})^{-1} \mathbf{y} &= \text{Vec}(\mathbf{Y})^T (\mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}) \text{Vec}(\mathbf{Y}) \\ &= \text{Vec}(\mathbf{Y})^T \text{Vec}(\mathbf{\Lambda}^{-1} \mathbf{Y} \mathbf{A}^{-1}) \\ &= \text{Tr}(\mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y} \mathbf{A}^{-1}) \\ &= \text{Tr}(\mathbf{S}_2 \mathbf{A}^{-1}) \end{aligned} \quad (11)$$

where

$$\mathbf{S}_2 = \mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y} \quad (12)$$

is the desired $n \times n$ matrix. Further, note that

$$\text{Tr}(\mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y} \mathbf{A}^{-1}) = \text{Tr}(\mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T \mathbf{\Lambda}^{-1})$$

Hence,

$$\mathbf{y}^T (\mathbf{A} \otimes \mathbf{\Lambda})^{-1} \mathbf{y} = \text{Tr}(\mathbf{S}_1 \mathbf{\Lambda}^{-1}) \quad (13)$$

where

$$\mathbf{S}_1 = \mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T \quad (14)$$

is the desired $m \times m$ matrix.

(d)

By part (b) and (c), the $m \times n$ random matrix \mathbf{Y} has distribution

$$\mathbf{Y} | \mathbf{A}, \mathbf{\Lambda} \sim \mathcal{MN}(\mathbf{O}, \mathbf{\Lambda}, \mathbf{A})$$

Note that here my notation of $\mathcal{MN}(\cdot, \cdot, \cdot)$ is different from what Dr. Banerjee demonstrated in class. Dear Dr. Banerjee switched the position of $\mathbf{\Lambda}$ and \mathbf{A} while the above is consistent with Wikipedia's definition (BTW, problem 3 uses the same definition as Wikipedia does). I choose to use Wikipedia's definition simply because we can check it with Wikipedia — each column of \mathbf{Y} is an $m \times 1$ vector of outcomes with covariance $\mathbf{\Lambda}$ while each row of \mathbf{Y} is a particular outcome with n different measurements, and the covariance is \mathbf{A} . The density of \mathbf{Y} given both \mathbf{A} and $\mathbf{\Lambda}$ is proportional to

$$|\mathbf{A}|^{-\frac{m}{2}} |\mathbf{\Lambda}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr}(\mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y} \mathbf{A}^{-1}) \right\}$$

Next, though not explicitly stated in problem, I assume that the joint prior on $(\mathbf{A}, \mathbf{\Lambda})$ are seperable, i.e., \mathbf{A} and $\mathbf{\Lambda}$ are independent so that

$$p(\mathbf{A} | \mathbf{\Lambda}) =_{a.s.} p(\mathbf{A})$$

and

$$p(\mathbf{\Lambda} | \mathbf{A}) =_{a.s.} p(\mathbf{\Lambda})$$

Hence, the conditional posterior for \mathbf{A} given \mathbf{Y} and $\mathbf{\Lambda}$ is

$$\begin{aligned}
p(\mathbf{A}|\mathbf{y}, \mathbf{\Lambda}) &= p(\mathbf{A}|\mathbf{Y}, \mathbf{\Lambda}) \\
&= \mathcal{MN}(\mathbf{Y}|\mathbf{A}, \mathbf{\Lambda}) \times \mathcal{IW}(\mathbf{A}|\nu_A, \mathbf{S}_A) \\
&\propto |\mathbf{A}|^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} \text{Tr}(\mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y} \mathbf{A}^{-1}) \right\} \times \\
&\quad |\mathbf{A}|^{-\frac{\nu_A + n + 1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr}(\mathbf{S}_A \mathbf{A}^{-1}) \right\} \\
&\propto |\mathbf{A}|^{-\frac{\nu_A + m + n + 1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr}([\mathbf{S}_A + \mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y}] \mathbf{A}^{-1}) \right\}
\end{aligned}$$

But this is just proportional to another Inverse Wishart density. Thus,

$$p(\mathbf{A}|\mathbf{y}, \mathbf{\Lambda}) = \mathcal{IW}(\nu_A + m, \mathbf{S}_A + \mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y}) \quad (15)$$

Similarly, switch ν_A to ν_Λ , \mathbf{S}_A to \mathbf{S}_Λ and $\mathbf{Y}^T \mathbf{\Lambda}^{-1} \mathbf{Y}$ to $\mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T$, we have

$$p(\mathbf{\Lambda}|\mathbf{y}, \mathbf{A}) = \mathcal{IW}(\nu_\Lambda + n, \mathbf{S}_\Lambda + \mathbf{Y} \mathbf{A}^{-1} \mathbf{Y}^T) \quad (16)$$

(Note: for some reason, I don't know why the right-hand-side does not become large.)

Problem 3

(a)

The dimensions of \mathbf{C} , \mathbf{V} , ν and \mathbf{S} are

Parameter	Dimension	Parameter	Dimension
\mathbf{C}	$p \times m$	\mathbf{V}	$p \times p$
ν	Scalar	\mathbf{S}	$m \times m$

(b)

By the usual technique,

$$\begin{aligned}
 p(\mathbf{B}, \boldsymbol{\Sigma} | \mathbf{Y}) &\propto p(\mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}) \\
 &= \mathcal{MN}(\mathbf{Y} | \mathbf{XB}, \mathbf{H}, \boldsymbol{\Sigma}) \times \mathcal{MN}\mathcal{IW}(\mathbf{B}, \boldsymbol{\Sigma} | \mathbf{C}, \mathbf{V}, \nu, \mathbf{S}) \\
 &= \mathcal{MN}(\mathbf{Y} | \mathbf{XB}, \mathbf{H}, \boldsymbol{\Sigma}) \times \mathcal{MN}(\mathbf{B} | \mathbf{C}, \mathbf{V}, \boldsymbol{\Sigma}) \times \mathcal{IW}(\boldsymbol{\Sigma} | \nu, \mathbf{S}) \\
 &\propto \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \exp \left\{ -\frac{1}{2} \text{Tr} ((\mathbf{Y} - \mathbf{XB})^T \mathbf{H}^{-1} (\mathbf{Y} - \mathbf{XB}) \boldsymbol{\Sigma}^{-1}) \right\} \times \\
 &\quad \frac{1}{|\boldsymbol{\Sigma}|^{p/2}} \exp \left\{ -\frac{1}{2} \text{Tr} ((\mathbf{B} - \mathbf{C})^T \mathbf{V}^{-1} (\mathbf{B} - \mathbf{C}) \boldsymbol{\Sigma}^{-1}) \right\} \times \\
 &\quad \frac{1}{|\boldsymbol{\Sigma}|^{(\nu+m+1)/2}} \exp \left\{ -\frac{1}{2} \text{Tr} (\mathbf{S} \boldsymbol{\Sigma}^{-1}) \right\} \\
 &= \frac{1}{|\boldsymbol{\Sigma}|^{p/2} |\boldsymbol{\Sigma}|^{(\nu+n+m+1)/2}} \exp \left\{ -\frac{1}{2} \text{Tr} (\boldsymbol{\Omega} \boldsymbol{\Sigma}^{-1}) \right\}
 \end{aligned}$$

Where $\boldsymbol{\Omega}$ is a matrix function of \mathbf{B} . Next, by Dr. Banerjee's Mm -trick (the difference here is that both M and m are matrices),

$$\begin{aligned}
 \boldsymbol{\Omega} &= \mathbf{Y}^T \mathbf{H}^{-1} \mathbf{Y} - \mathbf{Y}^T \mathbf{H}^{-1} \mathbf{XB} - \mathbf{B}^T \mathbf{X}^T \mathbf{H}^{-1} \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{H}^{-1} \mathbf{XB} + \\
 &\quad \mathbf{B}^T \mathbf{V}^{-1} \mathbf{B} - \mathbf{C}^T \mathbf{V}^{-1} \mathbf{B} - \mathbf{B}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{C}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{S} \\
 &= \mathbf{B}^T \underbrace{(\mathbf{X}^T \mathbf{H}^{-1} \mathbf{X} + \mathbf{V}^{-1})}_{\mathfrak{M}} \mathbf{B} - \underbrace{(\mathbf{Y}^T \mathbf{H}^{-1} \mathbf{X} + \mathbf{C}^T \mathbf{V}^{-1})}_{\mathfrak{m}^T} \mathbf{B} + \\
 &\quad \mathbf{B}^T \mathfrak{m} + \mathbf{Y}^T \mathbf{H}^{-1} \mathbf{Y} + \mathbf{C}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{S} \\
 &= (\mathbf{B} - \mathfrak{M} \mathfrak{m})^T \mathfrak{M}^{-1} (\mathbf{B} - \mathfrak{M} \mathfrak{m}) + \underbrace{\mathbf{Y}^T \mathbf{H}^{-1} \mathbf{Y} + \mathbf{C}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{S} - \mathfrak{m}^T \mathfrak{M} \mathfrak{m}}_{\mathfrak{c}}
 \end{aligned}$$

Therefore, the posterior density is proportional to

$$p(\mathbf{B}, \Sigma | \mathbf{Y}) \propto \frac{1}{|\Sigma|^{p/2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left((\mathbf{B} - \mathfrak{M}\mathbf{m})^T \mathfrak{M}^{-1} (\mathbf{B} - \mathfrak{M}\mathbf{m}) \Sigma^{-1} \right) \right\} \times \\ \frac{1}{|\Sigma|^{(\nu+n+m+1)/2}} \exp \left\{ -\frac{1}{2} \text{Tr} (\mathfrak{C} \Sigma^{-1}) \right\}$$

where

$$\begin{aligned} \mathfrak{M} &= \mathbf{X}^T \mathbf{H}^{-1} \mathbf{X} + \mathbf{V}^{-1} \\ \mathbf{m} &= \mathbf{V}^{-1} \mathbf{C} + \mathbf{X}^T \mathbf{H}^{-1} \mathbf{Y} \\ \mathfrak{C} &= \mathbf{Y}^T \mathbf{H}^{-1} \mathbf{Y} + \mathbf{C}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{S} - \mathbf{m}^T \mathfrak{M} \mathbf{m} \end{aligned}$$

Hence, we have

$$p(\mathbf{B} | \mathbf{Y}, \Sigma) = \mathcal{MN}(\mathbf{B} | \mathfrak{M}\mathbf{m}, \mathfrak{M}, \Sigma) \quad (17)$$

$$p(\Sigma | \mathbf{Y}, \mathbf{B}) = \mathcal{IW}(\Sigma | \nu + n, \mathfrak{C}) \quad (18)$$

$$p(\mathbf{B}, \Sigma | \mathbf{Y}) = \mathcal{MN IW}(\mathbf{B}, \Sigma | \mathfrak{M}\mathbf{m}, \mathfrak{M}, \nu + n, \mathfrak{C}) \quad (19)$$

(c)

As for me, I prefer to use Gibbs sampling. To make a connection with part (b), I will use **method of mixtures** or **composition sampling**.

1. For $t = 1, 2, \dots, T$, do:

- Sample Σ from $p(\Sigma | \mathbf{Y}, \mathbf{B})$. Call it $\Sigma^{(t)}$.
- Sample \mathbf{B} from $p(\mathbf{B} | \mathbf{Y}, \Sigma)$. Call it $\mathbf{B}^{(t)}$.

2. Output T -pairs of samples $\{(\mathbf{B}^{(t)}, \Sigma^{(t)}) : t = 1, 2, \dots, T\}$.

(d)

By the usual technique in multivariate Bayesian statistics,

$$\begin{aligned} p(\mathbf{Y}_* | \mathbf{Y}, \mathbf{B}, \Sigma) &\propto p(\mathbf{Y}_* | \mathbf{B}, \Sigma) \\ &= \mathcal{MN}(\tilde{\mathbf{Y}} | \tilde{\mathbf{X}}\mathbf{B}, \tilde{\mathbf{H}}, \Sigma) \end{aligned}$$

where

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}_* \end{pmatrix}, \quad \tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix}, \quad \tilde{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}$$

Next, Let $\mathbf{W} = \mathbf{Y} - \mathbf{X}\mathbf{B}$, $\mathbf{W}_* = (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B})$ and write out the density explicitly,

$$\begin{aligned} p(\mathbf{Y}_*|\mathbf{Y}, \mathbf{B}, \Sigma) &\propto \exp \left\{ -\frac{1}{2} \text{Tr} (\mathbf{W}_*^T \mathbf{H}_{21} \mathbf{W} + \mathbf{W}^T \mathbf{H}_{12} \mathbf{W}_* + \mathbf{W}_*^T \mathbf{H}_{22} \mathbf{W}_*) \Sigma^{-1} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{Tr} (\mathbf{Y}_*^T \mathfrak{M}^{-1} \mathbf{Y}_* - \mathbf{Y}_*^T \mathbf{m} - \mathbf{m}^T \mathbf{Y}_*) \Sigma^{-1} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{Tr} ((\mathbf{Y}_*^T - \mathfrak{M} \mathbf{m})^T \mathfrak{M}^{-1} (\mathbf{Y}_*^T - \mathfrak{M} \mathbf{m}) \Sigma^{-1}) \right\} \end{aligned}$$

where \mathfrak{M} and \mathbf{m} come from the extremely useful and famous Dr. Banerjee's Mm -trick:

$$\begin{aligned} \mathfrak{M} &= \mathbf{H}_{22}^{-1} \\ \mathbf{m} &= \mathbf{H}_{22} \mathbf{X}_* \mathbf{B} - \mathbf{H}_{21} \mathbf{W} \\ \mathfrak{M} \mathbf{m} &= \mathbf{X}_* \mathbf{B} - \mathbf{H}_{22}^{-1} \mathbf{H}_{21} \mathbf{W} \end{aligned}$$

Therefore, I derive the matrix-version of the conditional multivariate normal density in Biostat 200C and Biostat 250A:

$$p(\mathbf{Y}_*|\mathbf{Y}, \mathbf{B}, \Sigma) = \mathcal{MN}(\mathbf{Y}_* | \mathfrak{M} \mathbf{m}, \mathfrak{M}, \Sigma) \quad (20)$$

Compare it with the classical conditional multivariate normal formula:

$$\begin{aligned} \mathbb{E}(\mathbf{y}_*|\mathbf{y}, \tilde{\boldsymbol{\mu}}, \mathbf{H}) &= \boldsymbol{\mu}_* - \mathbf{H}_{22}^{-1} \mathbf{H}_{21} (\mathbf{y} - \boldsymbol{\mu}) \\ \text{Var}(\mathbf{y}_*|\mathbf{y}, \tilde{\boldsymbol{\mu}}, \mathbf{H}) &= \mathbf{H}_{22}^{-1} \end{aligned}$$

where $\mathbf{H} = \Sigma^{-1}$ is the precision matrix. Hence, two formulas match with each other perfectly!!!

Finally, I draw samples using **method of mixtures** or **composition sampling**.

1. For $t = 1, 2, \dots, T$, do:

- Sample (\mathbf{B}, Σ) from $p(\mathbf{B}, \Sigma|\mathbf{Y})$ in part (b). Call it $(\mathbf{B}^{(t)}, \Sigma^{(t)})$.
- Sample \mathbf{Y}_* from $p(\mathbf{Y}_*|\mathbf{Y}, \mathbf{B}, \Sigma)$. Call it $\mathbf{Y}_*^{(t)}$.

2. Output T -arrays of samples $\{(\mathbf{Y}_*^{(t)}, \mathbf{B}^{(t)}, \Sigma^{(t)}) : t = 1, 2, \dots, T\}$.

3. The resulting $\mathbf{Y}_*^{(t)}$ has marginal distribution $p(\mathbf{Y}_*|\mathbf{Y})$, which is known as the entire predictive distribution according to Dr. Banerjee.

Problem 4

(a)

For each of the 7 nodes, their parent nodes are listed as follows.

Node	Parent(s)	Node	Parent(s)
y_1		y_2	y_1
y_3	y_1, y_2	y_4	y_1, y_2, y_3
y_5	y_2, y_3, y_4	y_6	y_1, y_4, y_5
y_7	y_1, y_2, y_6		

Hence, the joint density $p(\mathbf{y})$ can be written as

$$\begin{aligned}
 p(\mathbf{y}) = & p(y_1) \times p(y_2|y_1) \times p(y_3|y_1, y_2) \times \\
 & p(y_4|y_1, y_2, y_3) \times p(y_5|y_2, y_3, y_4) \times \\
 & p(y_6|y_1, y_4, y_5) \times p(y_7|y_1, y_2, y_6)
 \end{aligned} \tag{21}$$

(b)

By the linear model representation, each node y_i is a linear combination of others with an independent error term. Besides, if diagonal elements of \mathbf{A} are non-zero, then they could be subtracted by a diagonal matrix so that the linear combination of each node y_i does not involve itself. Hence,

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{22}$$

where all 1's come from the conditional density in part (a).

(c)

(I assume that \mathbf{D} is non-singular, otherwise it is not invertible.)

Short answer: The precision matrix \mathbf{Q} can be expressed as

$$\mathbf{Q} = \Sigma^{-1} = (\mathbf{I} - \mathbf{A})^T \mathbf{D}^{-1} (\mathbf{I} - \mathbf{A}) \tag{23}$$

with lower triangular elements

$$\mathbf{Q} = \begin{pmatrix} s_1 & & & & & & \\ -e_2 + e_3 + e_4 + e_7 & s_2 & & & & & \\ -e_3 + e_4 & e_3 + e_4 + e_5 & s_3 & & & & \\ -e_4 + e_6 & -e_4 + e_5 & -e_4 + e_5 & s_4 & & & \\ e_6 & -e_5 & -e_5 & -e_5 + e_6 & s_5 & & \\ -e_6 + e_7 & e_7 & 0 & -e_6 & -e_6 & s_6 & \\ -e_7 & -e_7 & 0 & 0 & 0 & -e_7 & s_7 \end{pmatrix} \quad (24)$$

where $e_i = 1/d_i$ and s_i 's are positive numbers depending on e_i 's.

The elements that are necessarily 0 are

$$\{q_{63}, q_{73}, q_{74}, q_{75}\} \cup \{q_{36}, q_{37}, q_{47}, q_{57}\} \quad (25)$$

where q_{ij} 's are the $(ij)^{th}$ element of \mathbf{Q} .

Long answer: Re-write the equation $\mathbf{y} = \mathbf{A}\mathbf{y} + \boldsymbol{\eta}$ as

$$(\mathbf{I} - \mathbf{A})\mathbf{y} = \boldsymbol{\eta}$$

I claim that $(\mathbf{I} - \mathbf{A})$ is invertible. This can be verified by solving the inverse matrix analytically, but here I provide a proof that is similar to Dr. Banerjee's proof in class.

Proof. For a square full rank matrix \mathbf{X} , the unique solution to the linear system

$$\mathbf{X}\mathbf{b} = \mathbf{0}$$

is indeed $\mathbf{b} = \mathbf{0}$. Substitute \mathbf{X} by $(\mathbf{I} - \mathbf{A})$, then by the joint density formula 21, we have $b_1 = 0$ where b_i is the i^{th} element of \mathbf{b} . The reason is that in formula 21, node y_1 does not have any parent node. Next, since the parent node of y_2 is just y_1 and $b_1 = 0$, we have $b_2 = 0$. For node y_3 , the parent nodes are y_1 and y_2 . But $b_1 = b_2 = 0$ implies that $b_3 = 0$. A similar argument indicates that $b_4 = 0$. For node y_5 , although the conditional density does not depend on y_1 , it does involve y_2, y_3, y_4 and nothing else. Hence, we have $b_5 = 0$. Next, the fact that parent nodes of y_6 are y_1, y_4 and y_5 implies $b_6 = 0$. Finally, we have $b_7 = 0$. In conclusion, $\mathbf{b} = \mathbf{0}$.

Therefore, the square matrix $(\mathbf{I} - \mathbf{A})$ is indeed non-singular and $(\mathbf{I} - \mathbf{A})^{-1}$ exists. \square

Knowing that $(\mathbf{I} - \mathbf{A})^{-1}$ exists and \mathbf{D} is a non-singular matrix, we have

$$\begin{aligned} \mathbf{y} &= (\mathbf{I} - \mathbf{A})^{-1} \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \\ \text{Cov}(\mathbf{y}) = \boldsymbol{\Sigma} &= (\mathbf{I} - \mathbf{A})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{A})^{-T} \\ \mathbf{Q} = \boldsymbol{\Sigma}^{-1} &= (\mathbf{I} - \mathbf{A})^T \mathbf{D}^{-1} (\mathbf{I} - \mathbf{A}) \end{aligned}$$

To identify the necessary 0 elements of \mathbf{Q} , we have to make use of part (d): y_i and y_j are conditionally independent if and only if $q_{ij} = 0$ provided the joint Gaussian assumption. Next, note that

$$\tilde{p}(y_i|\cdot) = p(y_i|\text{Pa}[i]) \prod_{\{j \in \text{Ch}[i]\}} p(y_j|y_{\text{Pa}[j]}) \quad (26)$$

where $\tilde{p}(y_i|\cdot)$ is the conditional density of y_i given all other nodes, $\text{Pa}[i]$ are the parent nodes of i and $\text{Ch}[i]$ are the child nodes of i . Therefore, by looking at formula 21 again, immediately we have

$$\begin{aligned} \tilde{p}(y_1|\cdot) &\propto f_1(y_1, y_2, y_3, y_4, y_5, y_6, y_7) \\ \tilde{p}(y_2|\cdot) &\propto f_2(y_1, y_2, y_3, y_4, y_5, y_6, y_7) \\ \tilde{p}(y_3|\cdot) &\propto f_3(y_1, y_2, y_3, y_4, y_5) \\ \tilde{p}(y_4|\cdot) &\propto f_4(y_1, y_2, y_3, y_4, y_5, y_6) \\ \tilde{p}(y_5|\cdot) &\propto f_5(y_1, y_2, y_3, y_4, y_5, y_6) \\ \tilde{p}(y_6|\cdot) &\propto f_6(y_1, y_2, y_4, y_5, y_6, y_7) \\ \tilde{p}(y_7|\cdot) &\propto f_7(y_1, y_2, y_6, y_7) \end{aligned}$$

where $f_i(\cdot)$ are measurable functions that are proportional to the conditional density $\tilde{p}(y_i|\cdot)$. Hence, by part (d), the following elements are necessarily 0:

$$\{q_{63}, q_{73}, q_{74}, q_{75}\} \cup \{q_{36}, q_{37}, q_{47}, q_{57}\} \quad (27)$$

(d)

The second part of (d) can be derived directly from part (c). The pairs of nodes that are conditionally independent given all the other nodes are:

$$\{(y_3, y_6), (y_3, y_7), (y_4, y_7), (y_5, y_7)\} \quad (28)$$

For the following, I will show the first part of (d).

Let k be the number of nodes and in our case, $k = 7$. The joint density of \mathbf{y} is proportional to

$$p(\mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} \right\}$$

The quadratic term $\mathbf{y}^T \mathbf{Q} \mathbf{y}$ can be expanded as

$$\begin{aligned}
\mathbf{y}^T \mathbf{Q} \mathbf{y} &= \sum_{i=1}^k \sum_{j=1}^k y_i y_j q_{ij} \\
&= \sum_{i=1}^n y_i^2 q_{ii} + \sum_{i \neq j} y_i y_j q_{ij} \\
&= \underbrace{\left(y_i^2 q_{ii} + 2y_i y_j q_{ij} + \sum_{l \neq i, j} y_i y_l q_{il} \right)}_{(*)} + \left(\sum_{l \neq i} y_l^2 q_{ll} + \sum_{j, l \neq i \text{ and } j \neq l} y_j y_l q_{jl} \right)
\end{aligned}$$

so that only the $(*)$ term involves y_i . Hence,

$$\tilde{p}(y_i | \cdot) \propto \exp \left\{ -\frac{1}{2} (*) \right\}$$

If $q_{ij} = 0$, then the node y_j does not appear in $(*)$. In other words, $(*)$ is independent of y_j provided that $q_{ij} = 0$, which indicates that $\tilde{p}(y_i | \cdot)$ does not depend on the value of y_j . But this is indeed one of the equivalent definitions of conditional independence between node y_i and node y_j . Hence,

$$y_i \perp\!\!\!\perp y_j | y_{-(i,j)} \iff q_{ij} = 0 \quad (29)$$

where $y_{-(i,j)}$ are the nodes except for y_i and y_j .

Problem 5

(a)

If \mathbf{Q} is symmetric, then

$$\mathbf{M}^{-1}(\mathbf{I} - \mathbf{C}) = (\mathbf{I} - \mathbf{C}^T)\mathbf{M}^{-1}$$

After re-arrangement of terms,

$$\mathbf{C}\mathbf{M} = \mathbf{M}\mathbf{C}^T$$

Hence,

$$c_{ij}m_j = m_i c_{ji} \quad (30)$$

One of the simplest choices could be

$$m_i = m_j = a, \quad c_{ij} = c_{ji} \quad (31)$$

(b)

Let $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ be the eigenvectors of $(\mathbf{I} - \mathbf{C})$ and $\lambda_1, \lambda_2, \dots, \lambda_n$ be the corresponding eigenvalues. Since all λ_i 's are positive, $(\mathbf{I} - \mathbf{C})$ is of full rank. Hence, for any vector $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{y}_i$$

for some $\alpha_1, \alpha_2, \dots, \alpha_n$. If $\mathbf{x} \neq \mathbf{0}$, then at least one α_i is not 0. Next, by the definition of eigenvectors and eigenvalues,

$$\begin{aligned} \mathbf{x}^T \mathbf{Q} \mathbf{x} &= \sum_{i=1}^n \alpha_i^2 \mathbf{y}_i^T \mathbf{M}^{-1} (\mathbf{I} - \mathbf{C}) \mathbf{y}_i \\ &= \sum_{i=1}^n \alpha_i^2 \lambda_i \mathbf{y}_i^T \mathbf{M}^{-1} \mathbf{y}_i \end{aligned}$$

WLOG, suppose $\alpha_i \neq 0$, then $\alpha_i^2 \lambda_i \mathbf{y}_i^T \mathbf{M}^{-1} \mathbf{y}_i$ because $\lambda_i > 0$ and diagonal elements of \mathbf{M}^{-1} are positive. Therefore,

$$\mathbf{x}^T \mathbf{Q} \mathbf{x} > 0 \quad \forall \mathbf{x} \neq \mathbf{0} \quad (32)$$

But this is the definition of positive-definiteness. Let the spectral decomposition of \mathbf{Q} be

$$\mathbf{Q} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T$$

Then the inverse of \mathbf{Q} is

$$\mathbf{Q}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T$$

and $\mathbf{\Lambda}^{-1}$ is a diagonal matrix with positive diagonal elements. For any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x} = (\mathbf{P}\mathbf{x})^T \mathbf{\Lambda}^{-1} (\mathbf{P}\mathbf{x}) > 0 \quad (33)$$

Hence, \mathbf{Q}^{-1} is also positive definite.

(c)

One shortcut is to write the conditional assumption as a linear model:

$$\begin{aligned} \mathbf{x} &= \mathbf{C}\mathbf{x} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(\mathbf{0}, \mathbf{M}) \end{aligned}$$

Then the result follows immediately. But I think this is not what Dr. Banerjee want us to do — He wants us to apply Brook's lemma as we did in HW8 [1]:

$$\frac{p(x_1, x_2, \dots, x_n)}{p(x_{10}, x_{20}, \dots, x_{n0})} = \prod_{i=1}^n \underbrace{\frac{p(x_i | x_{10}, \dots, x_{(i-1)0}, x_{i+1}, \dots, x_n)}{p(x_{i0} | x_{10}, \dots, x_{(i-1)0}, x_{i+1}, \dots, x_n)}}_{Q(x_i | x_{i0})}$$

where $p(\cdot)$ is the density function and x_{i0} is a realization of x_i .

WLOG, assume $x_{i0} = 0$ for all i , then for $i = 1, 2, \dots, n$,

$$\begin{aligned} Q(x_i | x_{i0}) &\propto \frac{\exp \left\{ -\frac{1}{2m_i} \left(x_i - \sum_{j=1}^{i-1} c_{ij} \times 0 - \sum_{j=i+1}^n c_{ij} x_j \right)^2 \right\}}{\exp \left\{ -\frac{1}{2m_i} \left(0 - \sum_{j=1}^{i-1} c_{ij} \times 0 - \sum_{j=i+1}^n c_{ij} x_j \right)^2 \right\}} \\ &= \exp \left\{ -\frac{1}{2m_i} \left(x_i^2 - 2x_i \sum_{j=i+1}^n c_{ij} x_j \right) \right\} \end{aligned}$$

Next, note that

$$\sum_{i=1}^n \frac{1}{m_i} \times x_i^2 = \mathbf{x}^T \mathbf{M}^{-1} \mathbf{x} \quad (34)$$

$$\sum_{i=1}^n \frac{1}{m_i} \left(2x_i \sum_{j=i+1}^n c_{ij} x_j \right) = \mathbf{x}^T \mathbf{C} \mathbf{M}^{-1} \mathbf{x} \quad (35)$$

where the second equality comes from the facts:

- Diagonal elements of \mathbf{C} are zeros.

- The matrix \mathbf{Q} is p.s.d. and hence $c_{ij}/m_i = c_{ji}/m_j$ or $\mathbf{C}\mathbf{M}^{-1}$ is symmetric.

Then we have

$$p(x_1, x_2, \dots, x_n) \propto \prod_{i=1}^n Q(x_i | x_{i0}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\mathbf{I} - \mathbf{C}) \mathbf{M}^{-1} \mathbf{x} \right\} \quad (36)$$

Therefore, the joint density of \mathbf{x} is multivariate normal with zero mean and covariance matrix $\mathbf{M}(\mathbf{I} - \mathbf{C})^{-T}$ or $(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}$.

(d)

Suppose there are two different pairs of matrices (\mathbf{M}, \mathbf{A}) and (\mathbf{S}, \mathbf{B}) such that

$$\Sigma = \mathbf{M}^{-1}\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}$$

and \mathbf{M}, \mathbf{S} are diagonal and \mathbf{A}, \mathbf{B} has diagonal elements equal to one. Then we have

$$\mathbf{M}^{-1}\mathbf{A} - \mathbf{S}^{-1}\mathbf{B} = \mathbf{M}^{-1}(\mathbf{A} - \mathbf{M}\mathbf{S}^{-1}\mathbf{B}) = \mathbf{O}$$

Thus,

$$\mathbf{A} = \mathbf{M}\mathbf{S}^{-1}\mathbf{B}$$

Focusing on the diagonal elements of both sides:

$$1 = a_{ii} = \frac{m_{ii}}{s_{ii}} \times b_{ii} = \frac{m_{ii}}{s_{ii}}$$

Hence, we have shown $m_{ii} = s_{ii}, i = 1, 2, \dots, n$, or equivalently,

$$\mathbf{M} = \mathbf{S} \quad (37)$$

Plug-in the equality to the above one, we have

$$\mathbf{A} = \mathbf{B} \quad (38)$$

Finally, the fact that $\mathbf{A} = \mathbf{B} = (\mathbf{I} - \mathbf{C})$ completes the proof.

(e)

For part (e), I would like to refer to Dr. Banerjee's book [2] and Mardia's paper [3].

Again, by Brook's lemma,

$$\frac{p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)}{p(\mathbf{x}_{10}, \mathbf{x}_{20}, \dots, \mathbf{x}_{n0})} = \prod_{i=1}^n \underbrace{\frac{p(\mathbf{x}_i | \mathbf{x}_{10}, \dots, \mathbf{x}_{(i-1)0}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n)}{p(\mathbf{x}_{i0} | \mathbf{x}_{10}, \dots, \mathbf{x}_{(i-1)0}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n)}}_{\Delta(\mathbf{x}_i | \mathbf{x}_{i0})}$$

where now \mathbf{x}_{i0} is a vector of zeros instead of a zero scalar. Hence,

$$\begin{aligned} \Delta(\mathbf{x}_i | \mathbf{x}_{i0}) &\propto \frac{\exp \left\{ -\frac{1}{2} \left(\mathbf{x}_i - \sum_{j=i+1}^n \mathbf{C}_{ij} \mathbf{x}_j \right)^T \Gamma_i^{-1} \left(\mathbf{x}_i - \sum_{j=i+1}^n \mathbf{C}_{ij} \mathbf{x}_j \right) \right\}}{\exp \left\{ -\frac{1}{2} \left(\mathbf{0} - \sum_{j=i+1}^n \mathbf{C}_{ij} \mathbf{x}_j \right)^T \Gamma_i^{-1} \left(\mathbf{0} - \sum_{j=i+1}^n \mathbf{C}_{ij} \mathbf{x}_j \right) \right\}} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\mathbf{x}_i^T \Gamma_i^{-1} \mathbf{x}_i - 2 \mathbf{x}_i^T \Gamma_i^{-1} \sum_{j=i+1}^n \mathbf{C}_{ij} \mathbf{x}_j \right) \right\} \end{aligned}$$

Next, note that the summation of the first and the second term are

$$\begin{aligned} \sum_{i=1}^n \mathbf{x}_i^T \Gamma_i^{-1} \mathbf{x}_i &= \mathbf{x}^T \mathbf{Block}(\Gamma_i^{-1}) \mathbf{x} \\ 2 \sum_{i=1}^n \mathbf{x}_i^T \Gamma_i^{-1} \left(\sum_{j=i+1}^n \mathbf{C}_{ij} \mathbf{x}_j \right) &= \mathbf{x}^T \mathbf{Block}(\Gamma_i^{-1} \mathbf{C}_{ij}) \mathbf{x} \end{aligned}$$

where $\mathbf{C}_{ii} = \mathbf{O}$ and $\mathbf{Block}(\cdot)$ is the block operator. The second summation assumes that

$$\Gamma_i^{-1} \mathbf{C}_{ij} = \mathbf{C}_{ij}^T \Gamma_i^{-1} \quad (39)$$

This is one of the conditions such that \mathbf{Q} is symmetric where

$$\mathbf{Q} = \mathbf{Block}(\Gamma_i^{-1} (\mathbf{1}_{\{i=j\}} \mathbf{I}_p - \mathbf{C}_{ij})) = \mathbf{Block}(\text{Diag}(\Gamma_i^{-1})) \mathbf{Block}(\mathbf{1}_{\{i=j\}} \mathbf{I}_p - \mathbf{C}_{ij}) \quad (40)$$

provided $\mathbf{Q} \succ \mathbf{O}$ and $\mathbf{1}_{\{i=j\}}$ is the indicator function. Hence,

$$p(\mathbf{x}) \propto \prod_{i=1}^n \Delta(\mathbf{x}_i | \mathbf{x}_{i0}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} \right\} \quad (41)$$

Finally, if

$$\Gamma_i = \Gamma_j = \Gamma \text{ and } \mathbf{C}_{ij} = -c_{ij} \mathbf{I}_p, \quad c_{ii} = 0, \quad c_{ij} = c_{ji} \quad \forall i, j \quad (42)$$

Then

$$\mathbf{Q} = \mathbf{C} \otimes \Gamma^{-1}, \quad \mathbf{C} = [c_{ij}]_{i,j=1}^n \quad (43)$$

References

- [1] Sudipto Banerjee and Elvis Cui. Lecture notes for Biostat 250C. UCLA Unpublished Private Handwritten Notes, 2021.
- [2] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. Hierarchical modeling and analysis for spatial data. CRC press, 2014.
- [3] KV Mardia. Multi-dimensional multivariate gaussian markov random fields with application to image processing. Journal of Multivariate Analysis, 24(2):265–284, 1988.