UCLA Biostatistics 250C

Final

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

Time: Take-Home

Instructor: Professor Sudipto Banerjee

1. (20 points) The inverse-Wishart density is a popular prior distribution for modeling unknown positive-definite matrices. For example, if $\mathbf{\Sigma}$ is a $p \times p$ covariance matrix, then the inverse-Wishart density is given by

$$IW(\mathbf{\Sigma} \,|\, \alpha, \mathbf{\Omega}) \propto \frac{1}{|\mathbf{\Sigma}|^{\frac{\alpha+p+1}{2}}} \times \exp\left(-\frac{1}{2}\mathrm{Tr}(\mathbf{\Omega}\mathbf{\Sigma}^{-1})\right) \;,$$

where α is a scalar parameter and $\mathbf{\Omega}$ is another $p \times p$ matrix parameter. Let $\mathbf{x}$ be a $p \times 1$ random vector and consider the following model

$$IW(\mathbf{\Sigma} \,|\, \alpha, \mathbf{\Omega}) \times N\left(\boldsymbol{\theta} \,|\, \boldsymbol{\mu}, \rho\mathbf{\Sigma}\right) \times N(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{\Sigma}) \;,$$

where α, $\mathbf{\Omega}$, $\boldsymbol{\mu}$ and ρ are known and fixed quantities (α and ρ are scalars).

(a) Find $p(\boldsymbol{\theta} \,|\, \mathbf{\Sigma}, \mathbf{x})$; identify the distribution and the parameters involved.

(b) Find $p(\mathbf{\Sigma} \,|\, \mathbf{x})$; identify the distribution and the parameters involved.

(c) Describe how you will draw samples from $p(\boldsymbol{\theta}, \mathbf{\Sigma} \,|\, \mathbf{x})$ using *multivariate Normal* and *Inverse Wishart* generators only.

(d) Explain clearly why the samples of $\boldsymbol{\theta}$ obtained in part (c) are from $p(\boldsymbol{\theta} \,|\, \mathbf{x})$.

Solution

(a) The joint density is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{x}) \propto \frac{1}{|\boldsymbol{\Sigma}|^{\frac{\alpha+p+1}{2}}} \times \exp\left(-\frac{1}{2}\mathrm{Tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1})\right) \times \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left(-\frac{1}{2\rho}(\boldsymbol{\theta}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right)$$
$$\times \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\theta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\theta})\right).$$

The conditional density $p(\boldsymbol{\theta} \,|\, \boldsymbol{\Sigma}, \mathbf{x})$ is proportional to the joint density so we keep track of terms involving $\boldsymbol{\theta}$ only. Therefore, standard algebraic manipulation (seen in class several times) yields

$$p(\boldsymbol{\theta} \,|\, \boldsymbol{\Sigma}, \mathbf{x}) \propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^\top \left(\frac{\rho}{1+\rho}\boldsymbol{\Sigma}\right)^{-1} \boldsymbol{\theta} - 2\boldsymbol{\theta}^\top \boldsymbol{\Sigma}^{-1}\left(\frac{\boldsymbol{\mu}}{\rho} + \mathbf{x}\right) + \text{const.}\right]\right)$$
$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \mathbf{Mm})^\top \mathbf{M}^{-1}(\boldsymbol{\theta} - \mathbf{Mm})\right),$$

where $\mathbf{M} = \dfrac{\rho}{1+\rho}\boldsymbol{\Sigma}$ and $\mathbf{m} = \boldsymbol{\Sigma}^{-1}\left(\dfrac{\boldsymbol{\mu}}{\rho} + \mathbf{x}\right)$. Hence, $p(\boldsymbol{\theta} \,|\, \boldsymbol{\Sigma}, \mathbf{x}) = N(\boldsymbol{\theta} \,|\, \mathbf{Mm}, \mathbf{M})$.

(b) Since the joint density can be factorized as $p(\boldsymbol{\Sigma} \,|\, \mathbf{x}) \times p(\boldsymbol{\theta} \,|\, \boldsymbol{\Sigma}, \mathbf{x})$ and we have already obtained the form of $p(\boldsymbol{\theta} \,|\, \boldsymbol{\Sigma}, \mathbf{x})$ in (a), we can derive the factorization up to a proportionality constant. Specifically, note that the "constant" that gets absorbed into the proportionality constant in the last expression of (a) is explicitly given by

$$\frac{1}{\rho}\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{m}^\top \mathbf{Mm} = \mathrm{Tr}\left(\left[\frac{\boldsymbol{\mu}\boldsymbol{\mu}^\top}{\rho} + \mathbf{x}\mathbf{x}^\top - \frac{\rho}{1+\rho}\left(\mathbf{x} + \frac{\boldsymbol{\mu}}{\rho}\right)\left(\mathbf{x} + \frac{\boldsymbol{\mu}}{\rho}\right)^\top\right]\boldsymbol{\Sigma}^{-1}\right)$$
$$= \mathrm{Tr}\left(\left[\frac{1}{1+\rho}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\right]\boldsymbol{\Sigma}^{-1}\right)$$

where $\mathbf{M}$ and $\mathbf{m}$ are as derived in (a). Therefore, we can factorize the joint density as

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{x}) \propto \frac{1}{|\boldsymbol{\Sigma}|^{\frac{\alpha+p+2}{2}}} \times \exp\left(-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Omega}^*\boldsymbol{\Sigma}^{-1}\right)\right) \times N(\boldsymbol{\theta} \,|\, \mathbf{Mm}, \mathbf{M})$$
$$\propto IW\left(\boldsymbol{\Sigma} \,|\, \alpha+1, \boldsymbol{\Omega}^*\right) \times N(\boldsymbol{\theta} \,|\, \mathbf{Mm}, \mathbf{M}),$$

where $\boldsymbol{\Omega}^* = \boldsymbol{\Omega} + \dfrac{1}{1+\rho}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top$.

(c) Use method of mixtures: Draw $\boldsymbol{\Sigma} \sim IW(\alpha+1, \boldsymbol{\Omega}^*)$ and for each drawn value of $\boldsymbol{\Sigma}$ draw $\boldsymbol{\theta} \sim N(\mathbf{Mm}, \mathbf{M})$ using the drawn value of $\boldsymbol{\Sigma}$ to compute $\mathbf{M}$ and $\mathbf{m}$. This results in a sample of $\{\boldsymbol{\theta}, \boldsymbol{\Sigma}\} \sim p(\boldsymbol{\theta}, \boldsymbol{\Sigma} \,|\, \mathbf{x})$.

(d) This has been proved in class. Check class notes.

2. (20 points) Let $\mathbf{y}_i$ be an $m \times 1$ vector of outcomes such that

$$\mathbf{y}_i \sim N(\mathbf{0}, a_{ii}\mathbf{\Lambda}) \text{ and } \operatorname{cov}\{\mathbf{y}_i, \mathbf{y}_j\} = a_{ij}\mathbf{\Lambda} \quad \text{for} \quad i, j = 1, 2, \ldots, n\,,$$

where $\mathbf{\Lambda}$ is an unknown $m \times m$ symmetric positive definite matrix and a_{ij}'s are unknown scalars such that the $n \times n$ matrix with a_{ij} as its (i, j)-th element, say $\mathbf{A} = \{a_{ij}\}$, is an $n \times n$ symmetric positive definite matrix. Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ be the $mn \times 1$ vector formed by stacking up the $\mathbf{y}_i$'s in a single column.

(a) Show that the variance-covariance matrix of $\mathbf{y}$, conditional upon the parameters, is $\operatorname{var}\{\mathbf{y}\} = \mathbf{A} \otimes \mathbf{\Lambda}$, where $\otimes$ is the *Kronecker product* for matrices.

(b) Show that there exists an $m \times m$ matrix $\mathbf{S}_1$ such that

$$\mathbf{y}^\top (\mathbf{A} \otimes \mathbf{\Lambda})^{-1} \mathbf{y} = \operatorname{Tr}(\mathbf{S}_1 \mathbf{\Lambda}^{-1}),$$

where $\operatorname{Tr}(\cdot)$ denotes the trace function for a matrix.

(c) Show that there exists an $n \times n$ matrix $\mathbf{S}_2$ such that

$$\mathbf{y}^\top (\mathbf{A} \otimes \mathbf{\Lambda})^{-1} \mathbf{y} = \operatorname{Tr}(\mathbf{S}_2 \mathbf{A}^{-1})\,.$$

(d) Let $\mathbf{A} \sim IW(\nu_A, \mathbf{S}_A)$ and $\mathbf{\Lambda} \sim IW(\nu_\Lambda, \mathbf{S}_\Lambda)$, where the hyperparameters ν_A, ν_Λ, $\mathbf{S}_A$ and $\mathbf{S}_\Lambda$ are known. Find $p(\mathbf{A} \,|\, \mathbf{y}, \mathbf{\Lambda})$ and $p(\mathbf{\Lambda} \,|\, \mathbf{A}, \mathbf{y})$.

Solution

(a) The covariance matrix of $\mathbf{y}$, by definition, is the $mn \times mn$ block matrix whose (i, j)-th block is the $m \times m$ matrix $\text{cov}\{\mathbf{y}_i, \mathbf{y}_j\} = a_{ij}\mathbf{\Lambda}$. The definition of the Kronecker product immediately yields the covariance matrix of $\mathbf{y}$ to be $\mathbf{A} \otimes \mathbf{\Lambda}$.

(b) Let $\mathbf{Y} = [\mathbf{y}_1 : \mathbf{y}_2 : \ldots : \mathbf{y}_n]$ be the $m \times n$ matrix with $\mathbf{y}_i$ as its columns. Using the relations between $\text{vec}(\cdot)$ and $\otimes$ derived in class notes and that $(\mathbf{A} \otimes \mathbf{\Lambda})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}$, we know that

$$\begin{aligned}
\mathbf{y}^\top (\mathbf{A} \otimes \mathbf{\Lambda})^{-1} \mathbf{y} = \mathbf{y}^\top \left(\mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}\right) \mathbf{y} &= \text{vec}(\mathbf{Y})^\top \left(\mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}\right) \text{vec}(\mathbf{Y}) \\
&= \text{vec}(\mathbf{Y})^\top \text{vec}(\mathbf{X}) \ , \ \text{where } \mathbf{X} = \mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{A}^{-1} \ , \\
&= \text{Tr}\left(\mathbf{X}\mathbf{Y}^\top\right) = \text{Tr}\left(\mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^\top\right) = \text{Tr}\left(\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^\top\mathbf{\Lambda}^{-1}\right) \\
&= \text{Tr}\left(\mathbf{S}_1\mathbf{\Lambda}^{-1}\right) \ , \ \text{where } \mathbf{S}_1 = \mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^\top \ .
\end{aligned}$$

Since $\mathbf{Y}$ is $m \times n$, it follows that $\mathbf{S}_1$ is $m \times m$.

(c) This will be similar to (b). In fact, the first few steps are identical:

$$\begin{aligned}
\mathbf{y}^\top (\mathbf{A} \otimes \mathbf{\Lambda})^{-1} \mathbf{y} = \mathbf{y}^\top \left(\mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}\right) \mathbf{y} &= \text{vec}(\mathbf{Y})^\top \left(\mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}\right) \text{vec}(\mathbf{Y}) \\
&= \text{vec}(\mathbf{Y})^\top \text{vec}(\mathbf{X}) \ , \ \text{where } \mathbf{X} = \mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{A}^{-1} \ , \\
&= \text{Tr}\left(\mathbf{X}\mathbf{Y}^\top\right) = \text{Tr}\left(\mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{A}^{-1}\mathbf{Y}^\top\right) = \text{Tr}\left(\mathbf{Y}^\top\mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{A}^{-1}\right) \\
&= \text{Tr}\left(\mathbf{S}_2\mathbf{A}^{-1}\right) \ , \ \text{where } \mathbf{S}_2 = \mathbf{Y}^\top\mathbf{\Lambda}^{-1}\mathbf{Y} \ .
\end{aligned}$$

Since $\mathbf{Y}^\top$ is $n \times m$, it follows that $\mathbf{S}_2$ is $n \times n$.

(d) The density $p(\mathbf{A} \,|\, \mathbf{y}, \mathbf{\Lambda})$ is proportional to

$$\begin{aligned}
p(\mathbf{A} \,|\, \mathbf{y}, \mathbf{\Lambda}) &\propto p(\mathbf{A}) \times p(\mathbf{y} \,|\, \mathbf{0}, \mathbf{A} \otimes \mathbf{\Lambda}) = IW(\mathbf{A} \,|\, \nu_A, \mathbf{S}_A) \times N(\mathbf{y} \,|\, \mathbf{0}, \mathbf{A} \otimes \mathbf{\Lambda}) \\
&\propto \frac{1}{|\mathbf{A}|^{\frac{\nu_A+n+1}{2}}} \exp\left(-\frac{1}{2}\text{Tr}\left(\mathbf{S}_A\mathbf{A}^{-1}\right)\right) \times \frac{1}{|\mathbf{A}|^{m/2}} \exp\left(-\frac{1}{2}\mathbf{y}^\top \left(\mathbf{A}^{-1} \otimes \mathbf{\Lambda}^{-1}\right) \mathbf{y})\right) \\
&\propto \frac{1}{|\mathbf{A}|^{\frac{nu_A+m+n+1}{2}}} \times \exp\left(-\frac{1}{2}\text{Tr}\left(\mathbf{S}_A\mathbf{A}^{-1}\right)\right) \times \exp\left(-\frac{1}{2}\text{Tr}\left(\mathbf{S}_2\mathbf{A}^{-1}\right)\right) \\
&\propto \frac{1}{|\mathbf{A}|^{\frac{nu_A+m+n+1}{2}}} \times \exp\left(-\frac{1}{2}\text{Tr}\left(\mathbf{S}_A\mathbf{A}^{-1} + \mathbf{S}_2\mathbf{A}^{-1}\right)\right) \\
&= IW\left(\mathbf{A} \,|\, \nu_A + m, \mathbf{S}_A + \mathbf{S}_2\right) \ ,
\end{aligned}$$

where $\mathbf{S}_2$ is from (c). In exactly analogous manner we find that $p(\mathbf{\Lambda} \,|\, \mathbf{y}, \mathbf{A}) = IW\left(\nu_\Lambda + n, \mathbf{S}_\Lambda + \mathbf{S}_1\right)$, where $\mathbf{S}_1$ is from (b).

3. (20 points) In a designed epidemiological experiment, measurements are collected on m different diseases over n different counties of California. The incidence rates for disease j in county i are indicated by y_{ij} and arranged as the (i,j)-th element in the $n \times m$ matrix $\mathbf{Y}$. Let $\mathbf{X}$ be a fixed $n \times p$ design matrix (to account for risk factors and potential confounders) with linearly independent columns (and $n > p$). Let $\mathbf{B}$ be a $p \times m$ matrix of the unknown regression coefficients. It is assumed that $\mathbf{Y}$ follows a Matrix-Normal distribution and the following linear regression model is constructed:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}\ ; \quad \mathbf{E} \sim MN(\mathbf{O}, \mathbf{H}, \mathbf{\Sigma})\ ,$$

where $MN(\mathbf{O}, \mathbf{H}, \mathbf{\Sigma})$ denotes the matrix normal family with $\mathbf{O}$ representing $n \times m$ matrix of zeroes for the mean of $\mathbf{E}$, $\mathbf{H}$ is an $n \times n$ spatial covariance matrix and $\mathbf{\Sigma}$ is the $m \times m$ covariance matrix representing covariances among the m diseases. The matrix $\mathbf{H}$ is positive-definite and is constructed from spatial methods. It is assumed fixed and known for this example. The density is given by

$$p(\mathbf{E}\,|\,\mathbf{H}, \mathbf{\Sigma}) = MN(\mathbf{E}\,|\,\mathbf{O}, \mathbf{H}, \mathbf{\Sigma}) = \frac{\exp\left[\mathrm{Tr}(\mathbf{E}^\top \mathbf{H}^{-1}\mathbf{E}\mathbf{\Sigma}^{-1})\right]}{(2\pi)^{mn/2}(\det(\mathbf{H}))^{m/2}(\det(\mathbf{\Sigma}))^{n/2}}\ .$$

It is of interest to estimate the regression coefficient slope matrix $\mathbf{B}$ and the covariance matrix among the diseases $\mathbf{\Sigma}$. For this, a Bayesian model is used (smart choice!) and the conjugate MNIW family is used to model $\{\mathbf{B}, \mathbf{\Sigma}\}$:

$$p(\mathbf{B}, \mathbf{\Sigma}) = MNIW(\mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{C}, \mathbf{V}, \nu, \mathbf{S}) = IW(\mathbf{\Sigma}\,|\,\nu, \mathbf{S}) \times MN(\mathbf{B}\,|\,\mathbf{C}, \mathbf{V}, \mathbf{\Sigma})\ ,$$

where $\mathbf{C}$, $\mathbf{V}$, ν and $\mathbf{S}$ are fixed hyperparameters.

(a) Write down the dimensions of $\mathbf{C}$, $\mathbf{V}$, ν and $\mathbf{S}$.

(b) Show that the posterior density $p(\mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{Y})$ follows an $MNIW$ distribution. Clearly write down the parameters of this $MNIW$ distribution.

(c) Explain clearly how you will sample $\{\mathbf{B}, \mathbf{\Sigma}\}$ from $p(\mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{Y})$.

(d) Your client is interested in "predicting" the disease rates for the m diseases in r additional counties where none of the disease rates were measured. The design matrix is available for these r counties, which we denote as the $r \times p$ matrix $\mathbf{X}_*$. Also, using Geographic Information systems (GIS) we calculate and fix the $(n+r) \times (n+r)$ spatial covariance matrix $\tilde{\mathbf{H}}$. Consider the following augmented model to accommodate predictions:

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{Y}_* \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_* \end{bmatrix} \mathbf{B} + \begin{bmatrix} \mathbf{E} \\ \mathbf{E}_* \end{bmatrix}\ ,$$

where $\mathbf{Y}_*$ is the $r \times m$ random matrix of unknown disease rates for the r counties and m diseases (formed analogous to the data matrix $\mathbf{Y}$). The $(n+r) \times m$ matrix $\tilde{\mathbf{E}} = \begin{bmatrix} \mathbf{E} \\ \mathbf{E}_* \end{bmatrix}$ is distributed as $\tilde{\mathbf{E}} \sim MN(\mathbf{O}, \tilde{\mathbf{H}}, \mathbf{\Sigma})$. Note that the joint posterior distribution of all "unknowns" given the data now includes the predictive matrix $\mathbf{Y}_*$ is

$$p(\mathbf{Y}_*, \mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{Y}) = p(\mathbf{Y}_*\,|\,\mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}) \times p(\mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{Y})\ .$$

You have already found $p(\mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{Y})$ as $MNIW$ in part (b). Can you find the distribution of $p(\mathbf{Y}_*\,|\,\mathbf{Y}, \mathbf{B}, \mathbf{\Sigma})$? Explain how you can draw samples from $p(\mathbf{Y}_*, \mathbf{B}, \mathbf{\Sigma}\,|\,\mathbf{Y})$ using the samples from part (c) and samples from $p(\mathbf{Y}_*\,|\,\mathbf{Y}, \mathbf{B}, \mathbf{\Sigma})$. Remark: The posterior samples of $\mathbf{Y}_*$ thus drawn gives the entire predictive distribution for the disease rate for each of the m diseases in each of the r unobserved counties.

Solution

(a) From the description of the problem, $\mathbf{C}$ is $p \times m$, $\mathbf{V}$ is $p \times p$, $\mathbf{S}$ is $m \times m$ and ν is 1×1 (scalar).

(b) The posterior density is

$$\begin{aligned} p(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y}) &\propto MNIW(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{C}, \mathbf{V}, \nu, \mathbf{S}) \times MN(\mathbf{Y} \,|\, \mathbf{XB}, \mathbf{H}, \mathbf{\Sigma}) \\ &\propto IW(\mathbf{\Sigma} \,|\, \nu, \mathbf{S}) \times MN(\mathbf{B} \,|\, \mathbf{C}, \mathbf{V}, \mathbf{\Sigma}) \times MN(\mathbf{Y} \,|\, \mathbf{XB}, \mathbf{H}, \mathbf{\Sigma}) \\ &\propto \frac{\exp\left[-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}\mathbf{\Sigma}^{-1}\right)\right]}{\det(\mathbf{\Sigma})^{\frac{\nu+m+1}{2}}} \times \frac{\exp\left[-\frac{1}{2}\mathrm{Tr}\left((\mathbf{B}-\mathbf{C})^\top \mathbf{V}^{-1}(\mathbf{B}-\mathbf{C})\mathbf{\Sigma}^{-1}\right)\right]}{\det(\mathbf{\Sigma})^{\frac{p}{2}}} \\ &\quad \times \frac{\exp\left[-\frac{1}{2}\mathrm{Tr}\left((\mathbf{Y}-\mathbf{XB})^\top \mathbf{H}^{-1}(\mathbf{Y}-\mathbf{XB})\mathbf{\Sigma}^{-1}\right)\right]}{\det(\mathbf{\Sigma})^{\frac{n}{2}}} . \end{aligned}$$

To simplify the above expression we recall the matrix-variate completion of squares:

$$\begin{aligned} &(\mathbf{B}-\mathbf{C})^\top \mathbf{V}^{-1}(\mathbf{B}-\mathbf{C}) + (\mathbf{Y}-\mathbf{XB})^\top \mathbf{H}^{-1}(\mathbf{Y}-\mathbf{XB}) \\ &= \mathbf{B}^\top \mathbf{M}^{-1}\mathbf{B} - \mathbf{B}^\top \mathbf{W} - \mathbf{W}^\top \mathbf{B} + \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} + \mathbf{Y}^\top \mathbf{H}^{-1}\mathbf{Y} \\ &= (\mathbf{B}-\mathbf{MW})^\top \mathbf{M}^{-1}(\mathbf{B}-\mathbf{MW}) + \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} + \mathbf{Y}^\top \mathbf{H}^{-1}\mathbf{Y} - \mathbf{W}^\top \mathbf{MW} , \\ &\text{where } \ \mathbf{M}^{-1} = \mathbf{V}^{-1} + \mathbf{X}^\top \mathbf{H}^{-1}\mathbf{X} \ \text{ and } \ \mathbf{W} = \mathbf{V}^{-1}\mathbf{C} + \mathbf{X}^\top \mathbf{H}^{-1}\mathbf{Y} . \end{aligned}$$

Using this to combine the last two exponents in the expression for $p(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y})$, we obtain

$$\begin{aligned} p(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y}) &\propto MNIW(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{C}, \mathbf{V}, \nu, \mathbf{S}) \times MN(\mathbf{Y} \,|\, \mathbf{XB}, \mathbf{H}, \mathbf{\Sigma}) \\ &\propto \frac{\exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbf{S}^*\mathbf{\Sigma}^{-1})\right]}{\det(\mathbf{\Sigma})^{\frac{\nu+n+m+1}{2}}} \times \frac{\exp\left[-\frac{1}{2}\mathrm{Tr}\left((\mathbf{B}-\mathbf{MW})^\top \mathbf{M}^{-1}(\mathbf{B}-\mathbf{MW})\mathbf{\Sigma}^{-1}\right)\right]}{(\det(\mathbf{\Sigma}))^{p/2}} \\ &\propto IW(\mathbf{\Sigma} \,|\, \nu^*, \mathbf{S}^*) \times MN(\mathbf{B} \,|\, \mathbf{MW}, \mathbf{M}, \mathbf{\Sigma}) = MNIW(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{MW}, \mathbf{M}, \nu^*, \mathbf{S}^*) , \end{aligned}$$

where $\nu^* = \nu + n$ and $\mathbf{S}^* = \mathbf{S} + \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} + \mathbf{Y}^\top \mathbf{H}^{-1}\mathbf{Y} - \mathbf{W}^\top \mathbf{MW}$.

(c) First draw $\mathbf{\Sigma} \sim IW(\nu^*, \mathbf{S}^*)$ and then, for this draw of $\mathbf{\Sigma}$, draw $\mathbf{B} \sim MN(\mathbf{MW}, \mathbf{M}, \mathbf{\Sigma})$.

(d) The predictive distribution is $p(\mathbf{Y}_*, \mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y}) = p(\mathbf{Y}_* \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}) \times p(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y})$. We have already found $p(\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{Y})$ in the previous part. To find $p(\mathbf{Y}_* \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma})$ we first note that

$$\begin{aligned} &\left[(\mathbf{Y}-\mathbf{XB})^\top : (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B})^\top\right] \begin{bmatrix} \tilde{\mathbf{H}}^{11} & \tilde{\mathbf{H}}^{12} \\ \tilde{\mathbf{H}}^{21} & \tilde{\mathbf{H}}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{Y}-\mathbf{XB} \\ \mathbf{Y}_* - \mathbf{X}_*\mathbf{B} \end{bmatrix} , \ \text{ where } \tilde{\mathbf{H}}^{-1} = \begin{bmatrix} \tilde{\mathbf{H}}^{11} & \tilde{\mathbf{H}}^{12} \\ \tilde{\mathbf{H}}^{21} & \tilde{\mathbf{H}}^{22} \end{bmatrix} ; \\ &\quad = (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B})^\top \tilde{\mathbf{H}}^{22}(\mathbf{Y}_* - \mathbf{X}_*\mathbf{B}) + (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B})^\top \mathbf{C} + \mathbf{C}^\top (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B}) + \text{terms free of } \mathbf{Y}_* , \end{aligned}$$

where $\mathbf{C} = \tilde{\mathbf{H}}^{21}(\mathbf{Y}-\mathbf{XB})$. Matrix-variate completion of squares now yields

$$\begin{aligned} &(\mathbf{Y}_* - \mathbf{X}_*\mathbf{B})^\top \tilde{\mathbf{H}}^{22}(\mathbf{Y}_* - \mathbf{X}_*\mathbf{B}) + (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B})^\top \mathbf{C} + \mathbf{C}^\top (\mathbf{Y}_* - \mathbf{X}_*\mathbf{B}) \\ &= \mathbf{Y}_*^\top \tilde{\mathbf{H}}^{22}\mathbf{Y}_* + \mathbf{Y}_*^\top \left[\mathbf{C} - \tilde{\mathbf{H}}^{22}\mathbf{X}_*\mathbf{B}\right] + \left[\mathbf{C} - \tilde{\mathbf{H}}^{22}\mathbf{X}_*\mathbf{B}\right]^\top \mathbf{Y}_* \\ &= (\mathbf{Y}_* - \mathbf{M}_*\mathbf{W}_*)^\top \mathbf{M}_*^{-1}(\mathbf{Y}_* - \mathbf{M}_*\mathbf{W}_*) + \text{terms free of } \mathbf{Y}_* , \end{aligned}$$

where $\mathbf{M}_*^{-1} = \tilde{\mathbf{H}}^{22}$ and $\mathbf{W}_* = \tilde{\mathbf{H}}^{22}\mathbf{X}_*\mathbf{B} - \mathbf{C}$. Hence, $\mathbf{M}_*\mathbf{W}_* = \mathbf{X}_*\mathbf{B} - \left(\tilde{\mathbf{H}}^{22}\right)^{-1}\tilde{\mathbf{H}}^{21}(\mathbf{Y} - \mathbf{XB})$ and $p(\mathbf{Y}_* \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{\Sigma}) = MN(\mathbf{Y}_* \,|\, \mathbf{M}_*\mathbf{W}_*, \mathbf{M}_*, \mathbf{\Sigma})$.

We can express this also in terms of the elements of $\tilde{\mathbf{H}}$ by recalling that $(\tilde{\mathbf{H}}^{22})^{-1} = \mathbf{F}$ and $\tilde{\mathbf{H}}^{21} = -\mathbf{F}^{-1}\tilde{\mathbf{H}}_{21}\tilde{\mathbf{H}}_{11}^{-1}$,where $\mathbf{F} = \tilde{\mathbf{H}}_{22} - \tilde{\mathbf{H}}_{21}\tilde{\mathbf{H}}_{11}^{-1}\tilde{\mathbf{H}}_{12}$. Therefore, $\mathbf{M}_* = \tilde{\mathbf{H}}_{22} - \tilde{\mathbf{H}}_{21}\tilde{\mathbf{H}}_{11}^{-1}\tilde{\mathbf{H}}_{12}$ and $\mathbf{M}_*\mathbf{W}_* = \mathbf{X}_*\mathbf{B} + \tilde{\mathbf{H}}_{21}\tilde{\mathbf{H}}_{11}^{-1}(\mathbf{Y} - \mathbf{XB})$, where $\tilde{\mathbf{H}}_{11} = \mathbf{H}$ in our problem.
Sampling: We draw one $\mathbf{Y}_* \sim MN(\mathbf{M}_*\mathbf{W}_*, \mathbf{M}_*, \mathbf{\Sigma})$ for every $\{\mathbf{B}, \mathbf{\Sigma}\}$ drawn in part (c).

4. (20 points) Bayesian networks (or directed acyclic graphs (DAG)) are increasingly being used for modeling joint distributions of a number of variables. Consider the following graph relating 7 health variables based upon certain conditional dependencies.

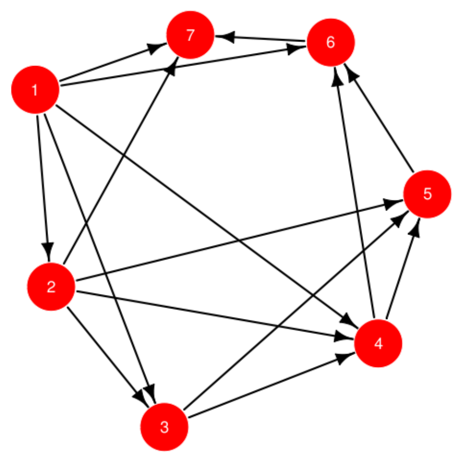

Figure 1: A Bayesian network of 7 variables with at most 3 parents (or "neighbors") for each node.

The 7 nodes are labeled by the integers $\{1, 2, \ldots, 7\}$ and correspond to a random vector $\mathbf{y} = [y_1, y_2, \ldots, y_7]^\top$, where y_i is the variable corresponding to node i for $i = 1, 2, \ldots, 7$. Answer the following questions based up the graph in Figure 1

(a) For each of the 7 nodes, list their parent nodes. Express the joint density $p(\mathbf{y})$ implied by the DAG.

(b) Suppose that the joint distribution $p(\mathbf{y})$ is multivariate Gaussian with a 7×1 mean vector $\boldsymbol{\mu}$ and a 7×7 variance-covariance matrix $\boldsymbol{\Sigma}$. Express the joint distribution of $\mathbf{y}$ in terms of the linear model,

$$\mathbf{y} = \mathbf{A}\mathbf{y} + \boldsymbol{\eta}\,, \quad \text{where} \quad \boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{D})\,,$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements d_i. Write down the structure of the 7×7 matrix $\mathbf{A}$ clearly indicating any elements that are necessarily 0 based upon Figure 1.

(c) Express the precision matrix $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ in terms of $\mathbf{A}$ and $\mathbf{D}$. What elements of $\mathbf{Q}$ are 0?

(d) Suppose $q_{ij} = 0$, where q_{ij} is the (i, j)-th element of $\mathbf{Q}$. Prove that $q_{ij} = 0$ implies that y_i is independent of y_j given the other elements in $\mathbf{y}$. For the DAG in Figure 1, find all pairs of nodes that are conditionally independent given all the other variables.

Solution

(a) The parents of the 7 nodes are read off from the graph:

$$\text{Pa}[1] = \{\phi\} \quad \text{empty set} \ ; \ \text{Pa}[2] = \{1\} \ ; \quad \text{Pa}[3] = \{1, 2\} \ ; \quad \text{Pa}[4] = \{1, 2, 3\} \ ;$$
$$\text{Pa}[5] = \{2, 3, 4\} \ ; \quad \text{Pa}[6] = \{1, 4, 5\} \ ; \quad \text{Pa}[7] = \{1, 2, 6\} \ .$$

The joint density is: $p(y_1) \times \prod_{i=2}^{7} p\left(y_i \,|\, y_{\text{Pa}[i]}\right)$.

(b) This will be represented as $y_1 = \eta_1 \sim N(0, d_1)$ and $y_i = \sum_{j=1}^{i-1} a_{ij} y_j + \eta_i$; $\eta_i \overset{ind}{\sim} N(0, d_i)$ for $i = 2, 3, \ldots, 7$, where, based upon the graph in Figure 1, $a_{ij} = 0$ for all $j \notin \text{Pa}[i]$. Writing this is matrix form, we obtain $\mathbf{y} = \mathbf{A}\mathbf{y} + \boldsymbol{\eta}$, where

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & 0 & 0 & 0 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & 0 & 0 & 0 & 0 \\ 0 & a_{52} & a_{53} & a_{54} & 0 & 0 & 0 \\ a_{61} & 0 & 0 & a_{64} & a_{65} & 0 & 0 \\ a_{71} & a_{72} & 0 & 0 & 0 & a_{76} & 0 \end{bmatrix}$$

and $\mathbf{D}$ is the diagonal matrix with elements d_i.

(c) Writing $\mathbf{y} = \mathbf{A}\mathbf{y} + \boldsymbol{\eta}$ as $(\mathbf{I} - \mathbf{A})\mathbf{y} = \boldsymbol{\eta}$, we can write $\mathbf{y} = (\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\eta}$. The inverse of $(\mathbf{I} - \mathbf{A})$ exists because $(\mathbf{I} - \mathbf{A})$ is a unit lower-triangular matrix (with 1's along the diagonal) and all triangular matrices with non-zero diagonals are invertible (easy linear algebra). From $\mathbf{y} = (\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\eta}$ it follows that $\boldsymbol{\Sigma} = \text{var}(\mathbf{y}) = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}(\mathbf{I} - \mathbf{A}^{\top})^{-1}$. Thus, $\mathbf{Q} = (\mathbf{I} - \mathbf{A}^{\top})\mathbf{D}^{-1}(\mathbf{I} - \mathbf{A})$. Getting the 0's in $\mathbf{Q}$ follows from direct matrix-multiplication. Alternatively, look at the joint density factorization in part (a) and note that the full conditional densities for each y_i will include only those factors that include y_i on the right hand side. Therefore, $p(y_i \,|\, \cdot) \propto p(y_i \,|\, y_{\text{Pa}[i]}) \times \prod_{j \neq i : i \in \text{Pa}[j]} p(y_j \,|\, y_{\text{Pa}[j]})$. We see that following elements (and their reflections along the diagonal) are zero: $\{(3, 6), (3, 7), (4, 7), (5, 7)\}$.

(d) The joint density of $p(\mathbf{y})$ is $p(\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{y}^{\top}\mathbf{Q}\mathbf{y}\right) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n} y_i \left(\sum_{j=1}^{n} q_{ij} y_j\right)\right)$.

From the above, the conditional density $p(y_i \,|\, \mathbf{y}_{-i})$, where $\mathbf{y}_{-i}$ is the vector formed by including all the elements of $\mathbf{y}$ except y_i, is given by

$$p(y_i \,|\, \mathbf{y}_{-i}) \propto \exp\left(-\frac{1}{2} y_i \left(\sum_{j=1}^{n} q_{ij} y_j\right)\right) \ .$$

Note that if $q_{ij} = 0$, then y_j drops out of the above exponent and there are no terms involving y_j in $p(y_i \,|\, \mathbf{y}_{-i})$. Therefore, if $q_{ij} = 0$, then $p(y_i \,|\, \mathbf{y}_{-i}) = p(y_i \,|\, \mathbf{y}_{-i,j})$ and, hence, y_i and y_j are conditionally independent given all other variables. Now, from the previous part, we obtain that all the zero elements of $\mathbf{Q}$ will indicate conditional independence given everything else. These will be: $\{y_3, y_6\}, \{y_3, y_7\}, \{y_4, y_7\}, \{y_5, y_7\}$.

5. (20 points) Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ be an $n \times 1$ vector of random variables. Suppose we construct the following set of conditional densities:

$$x_i \,|\, \mathbf{x}_{(-i)} \sim N\left(\sum_{j=1}^{n} c_{ij} x_j, m_i\right) \quad , \quad i = 1, 2, \ldots, n \; ,$$

where $\mathbf{x}_{(-i)}$ is the vector of all x_j with $j \neq i$, $c_{ii} = 0$ for all $i = 1, 2, \ldots, n$ and $m_i > 0$ for $i = 1, 2, \ldots, n$. Let $\mathbf{C}$ be the $n \times n$ matrix with elements c_{ij}, $\mathbf{M}$ be the $n \times n$ diagonal matrix with diagonal entries m_i for $i = 1, 2, \ldots, n$, and let $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$, where $\mathbf{I}$ is the $n \times n$ identity matrix. .

(a) Find a simple condition in terms of m_i, m_j, c_{ij} and c_{ji} so that $\mathbf{Q}$ is symmetric.

(b) Prove that $\mathbf{Q}$ is positive definite if and only if all eigenvalues of $\mathbf{I} - \mathbf{C}$ are positive. Explain why positive-definiteness of $\mathbf{Q}$ implies positive-definiteness of $\mathbf{Q}^{-1}$.

(c) Suppose $\mathbf{Q}$ is symmetric and positive definite. Prove that the joint density of $\mathbf{x}$ is multivariate normal with zero mean and covariance matrix $(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}$.

(d) Suppose you are given an $n \times n$ positive definite matrix $\mathbf{\Sigma}$. Show that there is a unique pair of matrices $\mathbf{C}$ and $\mathbf{M}$ such that $\mathbf{\Sigma}^{-1} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$, where the diagonal elements of $c_{ii} = 0$ and $\mathbf{M}$ is diagonal with diagonal entries $m_i > 0$.

(e) Consider the following multivariate extension: Let $\mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top, \ldots, \mathbf{x}_n^\top)^\top$ be an $np \times 1$ vector such that each component $\mathbf{x}_i$ is $p \times 1$. Consider the following conditional models:

$$\mathbf{x}_i \,|\, \mathbf{x}_{(-i)} \sim N\left(\sum_{j=1}^{n} \mathbf{C}_{ij} \mathbf{x}_j, \mathbf{\Gamma}_i\right) \quad , \quad i = 1, 2, \ldots, n \; ,$$

where $\mathbf{x}_{(-i)} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \setminus \{\mathbf{x}_i\}$, $\mathbf{C}_{ij}$'s are fixed $p \times p$ matrices, $\mathbf{C}_{ii} = \mathbf{O}$ (the matrix of zeroes), and $\mathbf{\Gamma}_i$'s are fixed positive definite matrices. Show that the joint density, if it exists, must be of the form:

$$p(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x}\right) \; .$$

Express the matrix $\mathbf{Q}$ in terms of the matrices $\mathbf{C}_{ij}$'s and $\mathbf{\Gamma}_i$'s and write down a condition in terms of $\mathbf{C}_{ij}$'s and $\mathbf{\Gamma}_i$'s for $\mathbf{Q}$ to be symmetric. Can you find specifications for each $\mathbf{C}_{ij}$ and $\mathbf{\Gamma}_i$ such that $\mathbf{Q}$ becomes a Kronecker product of two symmetric positive definite covariance matrices (neither of which is an identity matrix).

Solution

(a) Symmetry holds if and only if $\mathbf{M}^{-1}(\mathbf{I}-\mathbf{C}) = (\mathbf{I}-\mathbf{C}^\top)\mathbf{M}^{-1}$ or, equivalently, if and only if $\mathbf{M}^{-1}\mathbf{C} = \mathbf{C}^\top\mathbf{M}^{-1}$. This yields the necessary and sufficient condition: $c_{ij}/m_i = c_{ji}/m_j$.

(b) Note that $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I}-\mathbf{C}) = \mathbf{M}^{-1/2}(\mathbf{I}-\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2})\mathbf{M}^{-1/2}$. Therefore, $\mathbf{x}^\top\mathbf{Q}\mathbf{x} = \mathbf{y}^\top(\mathbf{I}-\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2})\mathbf{y}$, where $\mathbf{x} = \mathbf{M}^{1/2}\mathbf{y}$. Hence, the map between $\mathbf{x}$ and $\mathbf{y}$ is bijective (invertible). This implies that $\mathbf{Q}$ is positive definite if and only if $\mathbf{I}-\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2} = \mathbf{M}^{-1/2}(\mathbf{I}-\mathbf{C})\mathbf{M}^{1/2}$ is positive definite. Therefore, $\mathbf{I}-\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2}$ has the same eigenvalues as $\mathbf{I}-\mathbf{C}$ (since $\mathbf{I}-\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2}$ and $\mathbf{I}-\mathbf{C}$ are similar matrices), which implies that $\mathbf{Q}$ is positive definite if and only if all eigenvalues of $\mathbf{I}-\mathbf{C}$ are positive.

(c) We use Brook's lemma tp derive the joint distribution from its full conditionals:

$$\frac{p(\mathbf{x})}{p(\mathbf{x}_0)} = \prod_{i=1}^{n} \frac{p(x_i \,|\, x_{10},\ldots,x_{(i-1)0},x_{i+1},\ldots,x_n)}{p(x_{i0} \,|\, x_{10},\ldots,x_{(i-1)0},x_{i+1},\ldots,x_n)}\,,$$

where $\mathbf{x} = (x_1,x_2,\ldots,x_n)^\top$ and $\mathbf{x}_0 = (x_{10},x_{20},\ldots,x_{n0})^\top$. Setting $\mathbf{x}_0 = \mathbf{0}$ in the above expression yields

$$\begin{aligned}
p(\mathbf{x}) &\propto \prod_{i=1}^{n} \frac{\exp\left\{-\frac{1}{2m_i}\left(x_i - \sum_{j=i+1}^{n} c_{ij}x_j\right)^2\right\}}{\exp\left\{-\frac{1}{2m_i}\left(-\sum_{j=i+1}^{n} c_{ij}x_j\right)^2\right\}} \propto \prod_{i=1}^{n}\exp\left\{-\frac{1}{2m_i}\left(x_i^2 - 2x_i\sum_{j=i+1}^{n} c_{ij}x_j\right)\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n}\frac{x_i^2}{m_i} - 2\sum_{i=1}^{n} x_i\left(\frac{1}{m_i}\sum_{j=i+1} c_{ij}x_j\right)\right)\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n}\frac{x_i^2}{m_i} - 2\sum_{i=1}^{n}\sum_{j=i+1}\frac{c_{ij}}{m_i}x_ix_j\right)\right\} \propto \exp\left\{-\frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x}\right\}\,,
\end{aligned}$$

where $\mathbf{Q} = ((q_{ij}))$ with $q_{ii} = 1/m_i$ and $q_{ij} = -c_{ij}/m_i$; note that the quadratic form in the last expression makes use of the symmetry $c_{ij}/m_i = c_{ji}/m_j$ derived in (a) as we are given $\mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})$ is symmetric. Hence, $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})$ as defined in our problem.

(d) Write $\mathbf{\Sigma}^{-1} = \mathbf{M}^{-1} - (\mathbf{M}^{-1} - \mathbf{\Sigma}^{-1})$, where m_i is the the i-th diagonal element of $\mathbf{\Sigma}$. Solve for $\mathbf{C}$ in $\mathbf{M}^{-1}\mathbf{C} = \mathbf{M}^{-1} - \mathbf{\Sigma}^{-1}$ to produce $\mathbf{C} = \mathbf{I} - \mathbf{M}\mathbf{\Sigma}^{-1}$. This yields $\mathbf{\Sigma}^{-1} = \mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})$.

(e) Apply Brook's Lemma to the blocked vectors analogous to (c) to obtain the form:

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n}\mathbf{x}_i^\top\mathbf{\Gamma}_i^{-1}\mathbf{x}_i - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbf{x}_i^\top\mathbf{\Gamma}_i^{-1}\mathbf{C}_{ij}\mathbf{x}_j\right)\right\} \propto \exp\left(-\frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x}\right)\,,$$

where we must impose the condition that $\mathbf{\Gamma}_i^{-1}\mathbf{C}_{ij} = \mathbf{C}_{ji}^\top\mathbf{\Gamma}_j^{-1}$ or, equivalently, $\mathbf{C}_{ij}\mathbf{\Gamma}_j = \mathbf{\Gamma}_i\mathbf{C}_{ji}^\top$ so that the last expression is valid with $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I}_{np} - \mathbf{C})$ with $\mathbf{M} = \oplus_{i=1}^{n}\mathbf{\Gamma}_i$ (block diagonal with $\mathbf{\Gamma}_i$'s as $p\times p$ diagonal blocks) and $\mathbf{C} = [\mathbf{C}_{ij}]$ is the $np\times np$ matrix with $p\times p$ blocks $\mathbf{C}_{ij}$ for $i,j = 1,2,\ldots,n$. To get a Kronecker product, set $\mathbf{\Gamma}_i = \mathbf{\Gamma}$ for all $i = 1,2,\ldots,n$ and $\mathbf{C}_{ij} = c_{ij}\mathbf{I}_p$ and $c_{ii} = 0$ for $i,j = 1,2\ldots,n$. Then $\mathbf{Q} = \mathbf{C}\otimes\mathbf{\Gamma}^{-1}$.