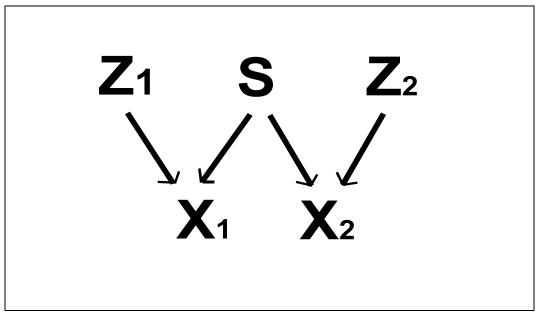
Stat 246 Homework IV - Solution

CA: Darongsae Kwon

1 [40 points]

(a) [5 points]



(b) [10 points]

Note that a set of random variables has a joint Gaussian distribution if and only if every linear combination of the random variables is normally distributed. Also note that independent normal random variables are jointly Gaussian. Since S, Z_1, Z_2 are independently normal, they are jointly Gaussian, and so every linear combination of them is normal. Since X_1 and X_2 are linear combinations of S, Z_1, Z_2, every linear combination of X_1 and X_2 is also a linear combination of S, Z_1, Z_2, and so normal. Thus, X_1 and X_2 are jointly Gaussian.

Note: Although each of the random variables in a set of jointly Gaussian random variables is necessarily Gaussian, it is possible for random variables to be individually Gaussian, but to not be jointly Gaussian.

$$
\begin{aligned}
&E(X_1) = E(S)a_1 + E(Z_1) = 0,\\
&E(X_2) = E(S)a_2 + E(Z_2) = 0,\\
&\mathrm{var}(X_1) = a_1^2\mathrm{var}(S) + \mathrm{var}(Z_1) = a_1^2\alpha^2 + \sigma^2,\\
&\mathrm{var}(X_2) = a_2^2\mathrm{var}(S) + \mathrm{var}(Z_2) = a_2^2\alpha^2 + \sigma^2,\\
&\mathrm{cov}(X_1, X_2) = a_1a_2\mathrm{var}(S) = a_1a_2\alpha^2.
\end{aligned}
$$

Thus,

$$\mu(\theta) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad C(\theta) = \begin{pmatrix} a_1^2\alpha^2 + \sigma^2 & a_1a_2\alpha^2 \\ a_1a_2\alpha^2 & a_2^2\alpha^2 + \sigma^2 \end{pmatrix}.$$

(c) [5 points]

Since $a_1^2 + a_2^2 = 1$,

$$\begin{pmatrix} a_1^2\alpha^2 + \sigma^2 & a_1a_2\alpha^2 \\ a_1a_2\alpha^2 & a_2^2\alpha^2 + \sigma^2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1^3\alpha^2 + a_1\sigma^2 + a_1a_2^2\alpha^2 \\ a_1^2a_2\alpha^2 + a_2^3\alpha^2 + a_2\sigma^2 \end{pmatrix} = (\alpha^2+\sigma^2)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix},$$

$$\begin{pmatrix} a_1^2\alpha^2 + \sigma^2 & a_1a_2\alpha^2 \\ a_1a_2\alpha^2 & a_2^2\alpha^2 + \sigma^2 \end{pmatrix}\begin{pmatrix} -a_2 \\ a_1 \end{pmatrix} = \begin{pmatrix} -a_1^2a_2\alpha^2 - a_2\sigma^2 + a_1^2a_2\alpha^2 \\ -a_1a_2^2\alpha^2 + a_1a_2^2\alpha^2 + a_1\sigma^2 \end{pmatrix} = \sigma^2\begin{pmatrix} -a_2 \\ a_1 \end{pmatrix}.$$

Therefore, $a = (a_1, a_2)$, $b = (-a_2, a_1)$ are eigenvectors of $C(\theta)$, and the corresponding eigenvalues are $\lambda = \alpha^2 + \sigma^2$, $\mu = \sigma^2$.

(d) [10 points]

Any symmetric positive definite matrix M has a spectral decomposition $M = UDU^t$, where D is a diagonal matrix whose diagonal entries are the eigenvalues of M and the columns of U are the corresponding orthonormal eigenvectors. Note that eigenvalues are unique and each eigenvector is unique up to sign.

Let C be a symmetric positive definite 2×2 matrix, with distinct eigenvalues $\lambda > \mu > 0$ and the unit length eigenvectors $a = (a_1, a_2)$ and $b = (-a_2, a_1)$. Then C, by the spectral decomposition, can be written as

$$C = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}\begin{pmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{pmatrix} = \begin{pmatrix} (\lambda-\mu)a_1^2 + \mu & (\lambda-\mu)a_1a_2 \\ (\lambda-\mu)a_1a_2 & (\lambda-\mu)a_2^2 + \mu \end{pmatrix}.$$

Let $\alpha^2 = \lambda - \mu$ and $\sigma^2 = \mu$, then $C = C(\theta)$ with $\theta = (a_1, a_2, \alpha^2, \sigma^2)$.

Now we find an invertible map between Σ^* and Θ. Define a map $\Phi : \Sigma^* \longrightarrow \Theta$ as follows. For $C \in \Sigma^*$ with distinct eigenvalues $\lambda > \mu > 0$, define $\Phi(C)$ as $\Phi(C) = (a_1, a_2, \alpha^2, \sigma^2) \in \Theta$ such that (a_1, a_2) is the unit length eigenvector corresponding to λ with $a_1 \geq 0$, $\alpha^2 = \lambda - \mu$ and $\sigma^2 = \mu$. As mentioned above, the spectral theorem says that $\lambda > \mu > 0$ are unique and $a = (a_1, a_2)$ is unique up to sign. Therefore, $\alpha^2 > 0$ and $\sigma^2 > 0$ are unique, and, due to the restriction $a_1 \geq 0$, (a_1, a_2) is also uniquely determined. This implies that Φ is well defined and one-to-one.

We can prove in addition that Φ is onto by showing that for each $\theta \in \Theta$, there exists some $C \in \Sigma^*$ such that $\Phi(C) = \theta$. For each $\theta = (a_1, a_2, \alpha^2, \sigma^2) \in \Theta$, let $C = UDU^t$, where U is an orthonormal matrix whose columns are $a = (a_1, a_2)$ and $b = (-a_2, a_1)$ and D is a diagonal matrix whose diagonal entries are $\lambda = \alpha^2 + \sigma^2$ and $\mu = \sigma^2$. Then the spectral theorem says that C has eigenvalues λ and μ. Since $\lambda > \mu > 0$, C is a symmetric positive definite matrix with two distinctive eigenvalues, so $C \in \Sigma^*$. Since $\alpha^2 = \lambda - \mu$, $\sigma^2 = \mu$, we have $\Phi(C) = \theta$. Therefore, Φ is one-to-one and onto, and so invertible.

(e) [10 points]

Note that Φ is invertible and denote its inverse by Φ^{-1}. Let $L(\theta; X_1, \ldots, X_N)$ be the likelihood as a function of θ and $L^*(C(\theta); X_1, \ldots, X_N) = L \circ \Phi(C(\theta); X_1, \ldots, X_N)$ be the likelihood as a function of $C(\theta) = \Phi^{-1}(\theta)$. Since we assume $\hat{C} \in \Sigma^*$, we have a unique $\hat{\theta} = \Phi(\hat{C})$ with $\hat{\theta} = (\hat{a}, \lambda_a - \lambda_b, \lambda_b)$. Given that $\hat{C}$ is the MLE of $C(\theta)$, we have

$$\max_{\theta} L(\theta; X_1, \ldots, X_N) = \max_{C(\theta)} L^*(C(\theta); X_1, \ldots, X_N) = L^*(\hat{C}) = L^*(\Phi^{-1}(\hat{\theta})) = L(\hat{\theta}).$$

Thus, $\hat{\theta}$ is the MLE of θ.

2 [20 points]

The joint distribution is given as $p(x) = \frac{1}{Z} \exp[\sum_{l=1}^{L} \psi_l(x_{C_l})]$.

Let $A, B, C \subset V$ are disjoint and C separates A from B. Suppose in addition that $A \cup B \cup C = V$. Since all paths from A to B go through C, we have the property that if $i \in A$ and $j \in B$, then i cannot be a neighbor of j, and so, they cannot be in the same clique.

Let

$$\begin{aligned} \mathcal{C} &= \{C_1, \ldots, C_L\}, \\ \mathcal{C}_A &= \{C_l \in \mathcal{C} : C_l \cap A \neq \emptyset\}, \\ \mathcal{C}_B &= \{C_l \in \mathcal{C} : C_l \cap B \neq \emptyset\}, \\ \mathcal{C}_* &= \mathcal{C} - (\mathcal{C}_A \cup \mathcal{C}_B). \end{aligned}$$

In other words, $\mathcal{C}_A$ is the set of maximal cliques that include at least one vertex in A, $\mathcal{C}_B$ is the set of maximal cliques that include at least one vertex in B, and $\mathcal{C}_*$ is the set of the remaining maximal cliques. Note that $\mathcal{C}_A \cup \mathcal{C}_B \cup \mathcal{C}_* = \mathcal{C}$ and $\mathcal{C}_A, \mathcal{C}_B, \mathcal{C}_*$ are disjoint (i.e., $\mathcal{C}_A \cup \mathcal{C}_B \cup \mathcal{C}_* = \mathcal{C}$ is a partition of $\mathcal{C}$).

Since

$$\begin{aligned} p(x) &= \frac{1}{Z} \exp[\sum_{l=1}^{L} \psi_l(x_{C_l})] \\ &= \frac{1}{Z} \exp[\sum_{l: C_l \in \mathcal{C}_A} \psi_l(x_{C_l})] \exp[\sum_{l: C_l \in \mathcal{C}_B} \psi_l(x_{C_l})] \exp[\sum_{l: C_l \in \mathcal{C}_*} \psi_l(x_{C_l})], \end{aligned}$$

we have

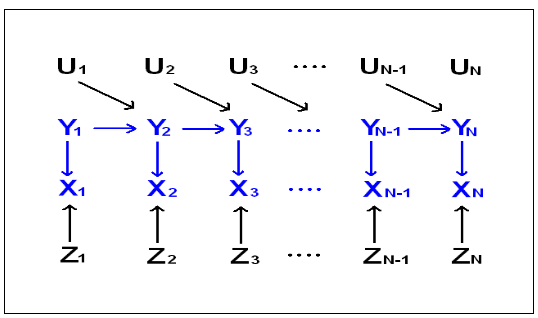
$$\begin{aligned}
p(x_B|x_A, x_C) &= \frac{p(x)}{p(x_A, x_C)} \\
&= \frac{p(x)}{\sum_{x_B} p(x)} \quad \text{(since } A \cup B \cup C = V\text{)} \\
&= \frac{\frac{1}{Z}\exp[\sum_{l:C_l\in\mathcal{C}_A}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_B}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_*}\psi_l(x_{C_l})]}{\sum_{x_B}\left\{\frac{1}{Z}\exp[\sum_{l:C_l\in\mathcal{C}_A}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_B}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_*}\psi_l(x_{C_l})]\right\}} \\
&= \frac{\frac{1}{Z}\exp[\sum_{l:C_l\in\mathcal{C}_A}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_B}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_*}\psi_l(x_{C_l})]}{\frac{1}{Z}\exp[\sum_{l:C_l\in\mathcal{C}_A}\psi_l(x_{C_l})]\exp[\sum_{l:C_l\in\mathcal{C}_*}\psi_l(x_{C_l})]\sum_{x_B}\exp[\sum_{l:C_l\in\mathcal{C}_B}\psi_l(x_{C_l})]} \\
&= \frac{\exp[\sum_{l:C_l\in\mathcal{C}_B}\psi_l(x_{C_l})]}{\sum_{x_B}\exp[\sum_{l:C_l\in\mathcal{C}_B}\psi_l(x_{C_l})]}
\end{aligned}$$

which doesn't depend on x_A (since cliques that contain some vertices in B don't contain any of the vertices in A) and is a function of x_B, x_C only. Therefore, $p(x_B|x_A, x_C) = p(x_B|x_C)$, or equivalently, $X_A \perp\!\!\!\perp X_B \mid X_C$.

To extend this result to the general case where $A \cup B \cup C \subset V$, let $\tilde{A} \supset A$ be all the nodes reached by paths from A before reaching C and $\tilde{B} = V - \tilde{A} - C \supset B$. Then $\tilde{A} \cup \tilde{B} \cup C = V$ are disjoint and C separates $\tilde{A}$ from $\tilde{B}$, so $X_{\tilde{A}} \perp\!\!\!\perp X_{\tilde{B}} \mid X_C$. Since X_A and X_B are subsets of $X_{\tilde{A}}$ and $X_{\tilde{B}}$, $X_A \perp\!\!\!\perp X_B \mid X_C$.

3 [40 points]

(a) [5 points]



Since $Y_{n+1} = Y_n + U_n$, we have $Y_{n+1} \perp\!\!\!\perp (Y_1, \ldots, Y_{n-1}) \mid Y_n$, so Y_n is a MC. If we only observe X_n, we can think that the MC Y_n is hidden but has a relationship with X_n such that $X_n = Y_n + Z_n$.

(b) [5 points]

Since Y_1, U_n, Z_n, $n = 1, \ldots, N$, are independent normal random variables, they are jointly Gaussian, and so every linear combination of them is normal. Every linear combination of all the variables is normal since every linear combination of all the variables is a linear combination Y_1, U_n, Z_n, $n = 1, \ldots, N$. Therefore, all the variables are jointly Gaussian. Since Y_1, U_n, Z_n, $n = 1, \ldots, N$ have mean 0, all the variables have mean 0.

(c), (d) [10 points]

Let $C_n = \begin{pmatrix} V_n & b_n \\ b_n^t & a_n \end{pmatrix}$ be the covariance matrix of $X_1, \ldots X_n, Y_n$, with V_n the covariance matrix of $(X_1, \ldots, X_n)$.

$$\begin{aligned}
\mathrm{cov}(Y_n, Y_{n+1}) &= \mathrm{cov}(Y_n, Y_n + U_n) \\
&= \mathrm{var}(Y_n) + \mathrm{cov}(Y_n, U_n) \\
&= \mathrm{var}(Y_1 + \sum_{k=1}^{n-1} U_k) \quad (\text{since } Y_n \perp\!\!\!\perp U_n) \\
&= \mathrm{var}(Y_1) + \sum_{k=1}^{n-1} \mathrm{var}(U_k) \\
&= \sigma^2 + (n-1)\sigma_u^2.
\end{aligned}$$

For $s = 1, \ldots, n$,

$$\begin{aligned}
\mathrm{cov}(X_s, Y_{n+1}) &= EE(X_s Y_{n+1} | Y_s) \\
&= E(E(X_s|Y_s)E(Y_{n+1}|Y_s)) \quad (\text{since } X_s \perp\!\!\!\perp Y_{n+1} \mid Y_s) \\
&= E(E(Y_s + Z_s|Y_s)E(Y_s + \sum_{k=s}^{n} U_k|Y_s)) \\
&= E(Y_s(Y_s + \sum_{k=s}^{n} E(U_k|Y_s))) \quad (\text{since } Z_s \perp\!\!\!\perp Y_s) \\
&= E(Y_s^2) \quad (\text{since } U_k \perp\!\!\!\perp Y_s, k \geq s) \\
&= \mathrm{var}(Y_s) = \sigma^2 + (s-1)\sigma_u^2.
\end{aligned}$$

$$\begin{aligned}
\mathrm{var}(Y_{n+1}) &= \mathrm{var}(Y_1 + \sum_{k=1}^{n} U_k) = \mathrm{var}(Y_1) + \sum_{k=1}^{n} \mathrm{var}(U_k) \\
&= \sigma^2 + n\sigma_u^2.
\end{aligned}$$

$$\begin{aligned}
\mathrm{var}(X_{n+1}) &= \mathrm{var}(Y_{n+1}) + \mathrm{var}(Z_{n+1}) + 2\mathrm{cov}(Y_{n+1}, Z_{n+1}) \\
&= \sigma^2 + n\sigma_u^2 + \sigma_z^2 + 2\mathrm{cov}(Y_1 + \sum_{k=1}^{n} U_k, Z_{n+1}) \\
&= \sigma^2 + n\sigma_u^2 + \sigma_z^2.
\end{aligned}$$

$$\begin{aligned}
\mathrm{cov}(X_{n+1}, Y_{n+1}) &= E(X_{n+1}Y_{n+1}) = EE(X_{n+1}Y_{n+1}|Y_{n+1}) \\
&= EE(Y_{n+1}^2 + Z_{n+1}Y_{n+1}|Y_{n+1}) \\
&= E(Y_{n+1}^2 + Y_{n+1}E(Z_{n+1}|Y_{n+1})) \\
&= \mathrm{var}(Y_{n+1}) = \sigma^2 + n\sigma_u^2. \quad (\text{since } Z_{n+1} \perp\!\!\!\perp Y_{n+1})
\end{aligned}$$

$$
\begin{aligned}
\mathrm{cov}(X_n, X_{n+1}) &= E(X_n X_{n+1}) = EE(X_n X_{n+1}|Y_n) \\
&= E(E(X_n|Y_n)E(X_{n+1}|Y_n)) \quad (\text{since } X_n \perp\!\!\!\perp X_{n+1} \mid Y_n) \\
&= E(Y_n(Y_n + E(U_n + Z_{n+1}|Y_n))) \quad (\text{since } U_n \perp\!\!\!\perp Y_n, Z_{n+1} \perp\!\!\!\perp Y_n) \\
&= \mathrm{var}(Y_n) = \sigma^2 + (n-1)\sigma_u^2.
\end{aligned}
$$

(e) [10 points]

For $s = 1, \ldots, n$,

$$
\begin{aligned}
\mathrm{cov}(X_s, X_{n+1}) &= \mathrm{cov}(X_s, Y_{n+1} + Z_{n+1}) \\
&= \mathrm{cov}(X_s, Y_{n+1}) + \mathrm{cov}(X_s, Z_{n+1}) \\
&= \mathrm{cov}(X_s, Y_{n+1}) \quad (\text{since } X_s \perp\!\!\!\perp Z_{n+1}) \\
&= \mathrm{var}(Y_s) = \sigma^2 + (s-1)\sigma_u^2.
\end{aligned}
$$

Thus, the $n+1$st column of V_{n+1} is $(\mathrm{var}(Y_1), \ldots, \mathrm{var}(Y_n), \mathrm{var}(X_{n+1}))$.

For $s = 1, \ldots, n-1$,

$$\mathrm{cov}(X_s, Y_n) = \mathrm{var}(Y_s) = \sigma^2 + (s-1)\sigma_u^2.$$

from (d).

Thus, $b_n = (\mathrm{var}(Y_1), \ldots, \mathrm{var}(Y_n))^t$. Then, $V_{n+1} = \begin{pmatrix} V_n & b_n \\ b_n^t & \mathrm{var}(X_{n+1}) \end{pmatrix}$.

From (d), $b_{n+1} = (\mathrm{var}(Y_1), \ldots, \mathrm{var}(Y_n), \mathrm{var}(Y_{n+1}))^t = (b_n^t, \mathrm{var}(Y_{n+1}))$, $\mathrm{var}(X_{n+1}) = \mathrm{var}(X_n) + \sigma_u^2$ and $\mathrm{var}(Y_{n+1}) = \mathrm{var}(Y_n) + \sigma_u^2$.

Thus,

$$
C_{n+1} = \begin{pmatrix} V_{n+1} & b_{n+1} \\ b_{n+1}^t & a_{n+1} \end{pmatrix} = \begin{pmatrix} V_n & b_n & b_n \\ b_n^t & \mathrm{var}(X_{n+1}) & \mathrm{var}(Y_{n+1}) \\ b_n^t & \mathrm{var}(Y_{n+1}) & \mathrm{var}(Y_{n+1}) \end{pmatrix} = \begin{pmatrix} V_n & b_n & b_n \\ b_n^t & V_n(n,n) + \sigma_u^2 & a_n + \sigma_u^2 \\ b_n^t & a_n + \sigma_u^2 & a_n + \sigma_u^2 \end{pmatrix},
$$

where $V_n(n,n)$ is the (n,n) entry of V_n.

(f) [10 points]

First, we have

$$
\begin{aligned}
E(Y_{n+1}|X_1, \ldots, X_n) &= E(Y_n + U_n|X_1, \ldots, X_n) \\
&= E(Y_n|X_1, \ldots, X_n) \quad (\text{since } U_n \perp\!\!\!\perp (X_1, \ldots, X_n)) \\
E(X_{n+1}|X_1, \ldots, X_n) &= E(Y_{n+1} + Z_{n+1}|X_1, \ldots, X_n) \\
&= E(Y_{n+1}|X_1, \ldots, X_n) \quad (\text{since } Z_{n+1} \perp\!\!\!\perp (X_1, \ldots, X_n)) \\
&= E(Y_n|X_1, \ldots, X_n).
\end{aligned}
$$

Since we know that the joint distribution of $(X_1, \ldots, X_n, Y_n)$ is Gaussian with mean zero and covariance C_n, we can compute $E(Y_n|X_1, \ldots, X_n)$.

If

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix}\right),$$

where Z_1 is a $p \times 1$ random vector and Z_2 is a $(n-p) \times 1$ random vector, then

$$\begin{aligned} E(Z_1 \mid Z_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2) \\ \mathrm{var}(Z_1 \mid Z_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^t. \end{aligned}$$

Thus, we have $E(Y_n \mid X_1, \ldots, X_n) = b_n^t V_n^{-1} X_{1:n}$, where $X_{1:n} = (X_1, \ldots, X_n)^t$. From this, we can get

$$E(Y_{n+1}|X_1, \ldots, X_n) = E(X_{n+1}|X_1, \ldots, X_n) = b_n^t V_n^{-1} X_{1:n}.$$