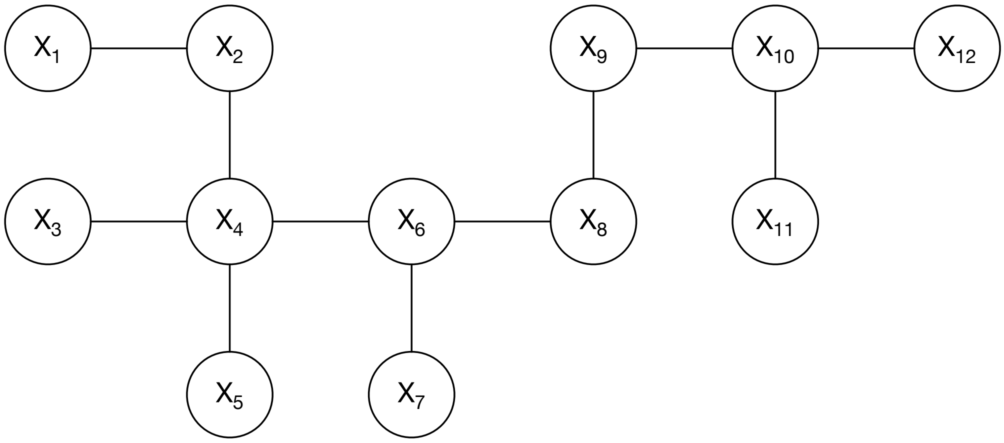
CS5340: Uncertainty Modeling in AI

Tutorial 5: Solutions

Released: Mar. 9, 2023

Problem 1. (MRT Inference)
We have previously considered inference for this MRF which models the activity (low or high) at 12 MRT stations. This week, we will repeat the activity but using the *sum product algorithm.*



Recall that each node represents a random variable indicating whether the activity at a particular station is low (0) or high (1) and assume the following factorization:

$$p(x_1, x_2, \ldots, x_{12}) = \frac{1}{Z} \prod_{i \in V} \psi(x_i) \prod_{(i,j) \in E} \psi(x_i, x_j) \tag{1}$$

where V is the set of nodes, E is the set of edges, and that the unary and pairwise factors are given by:

x_i	$\psi(x_i)$
0	10
1	2

Figure 1: Unary Factors

x_i	x_j	$\psi(x_i, x_j)$
0	0	20
0	1	5
1	0	5
1	1	20

Figure 2: Pairwise Factors

Note that the factors are the same across the nodes. Your task is to compute the following conditional probabilities using *the sum-product algorithm.*

Problem 1.a. [2 points] Compute $p(x_{12} = 1 | x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0)$.

Solution: Approach 1: According to the conditional independence in the MRF.

$$p(x_{12}|x_1, x_7, x_9, x_{10}) = p(x_{12}|x_{10}) = \frac{p(x_{10}, x_{12})}{\sum_{x_{12}} p(x_{10}, x_{12})}$$

Denote message from nodes $M_{E/x_{10},x_{12}} = \{x_1, \ldots x_9, x_{11}\}$ to x_{10} as

$$m(x_{10}) = \sum_{i,j \in M_{E/x_{10},x_{12}}} \psi(x_i)\psi(x_i, x_j)$$

Then

$$p(x_{12}|x_1, x_7, x_9, x_{10}) = \frac{p(x_{10}, x_{12})}{\sum_{x_{12}} p(x_{10}, x_{12})} = \frac{\sum_{i,j \in M_{E/x_{10},x_{12}}} p(x_1 \dots x_{12})}{\sum_{x_{12}} \sum_{i,j \in M_{E/x_{10},x_{12}}} p(x_1 \dots x_{12})}$$
$$= \frac{m(x_{10})\psi(x_{10})\psi(x_{10}, x_{12})\psi(x_{12})}{\sum_{x_{12}} m(x_{10})\psi(x_{10})\psi(x_{10}, x_{12})\psi(x_{12})}$$
$$= \frac{\psi(x_{10}, x_{12})\psi(x_{12})}{\sum_{x_{12}} \psi(x_{10}, x_{12})\psi(x_{12})} = \frac{\psi(x_{10} = 0, x_{12} = 1)\psi(x_{12} = 1)}{\sum_{x_{12}} \psi(x_{10} = 0, x_{12})\psi(x_{12})}$$
$$= \frac{5 \times 2}{20 \times 10 + 5 \times 2} = \frac{1}{21} = 0.0476$$

Approach 2: In the following solutions, we will compute the messages as follows.

$$m_{i \to j}(x_j) = \sum_{x_i \in \{0,1\}} \left(\psi^E(x_i)\psi(x_i, x_j) \prod_{x_k \in \mathrm{neighbors(x_i)} \backslash x_j} m_{k \to i}(x_i) \right) \tag{2}$$

where E is the set of evidence nodes, $\psi^E(x_i) = \delta(x_i = \hat{x}_i)\psi(x_i)$ if $x_i \in E$ and $\psi^E(x_i) = \psi(x_i)$ otherwise.

Node x_{12} is conditionally independent of all other nodes, given x_{10}. Let's compute the message from x_{10} to x_{12}.

$m_{x_{10} \to x_{12}}$		
$x_{12} = 0$	$10 \times 20 + 0 \times 5$	200
$x_{12} = 1$	$10 \times 5 + 0 \times 20$	50

$$\tilde{p}(x_{12} = \hat{x}_{12}|x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0) = \psi(x_{12} = \hat{x}_{12}) \times m_{x_{10} \to x_{12}}(x_{12} = \hat{x}_{12}) \tag{3}$$

$\tilde{p}$		
$x_{12} = 0$	10×200	2000
$x_{12} = 1$	2×50	100

$$p(x_{12} = 1|x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0) = \frac{100}{2000 + 100} = \frac{1}{21} = 0.0476 \tag{4}$$

Problem 1.b. [2 points] Compute $p(x_1 = 1|x_3 = 0, x_4 = 1, x_6 = 0)$.

Solution: Approach 1: According to the conditional independence in MRF.

$$p(x_1|x_3, x_4, x_6) = p(x_1|x_4) = \frac{p(x_1, x_4)}{\sum_{x_1} p(x_1, x_4)}$$

Denote message from nodes $M_{E/x_1,x_2,x_4} = \{x_3, x_5 \dots x_{12}\}$ to x_4 as

$$m(x_4) = \sum_{i,j \in M_{E/x_1,x_2,x_4}} \psi(x_i)\psi(x_i, x_j)$$

Then

$$p(x_1|x_3,x_4,x_6) = \frac{\sum_{x_2}\sum_{i,j\in M_{E/x_1,x_2,x_4}} p(x_1\dots x_{12})}{\sum_{x_2}\sum_{x_1}\sum_{i,j\in M_{E/x_1,x_2,x_4}} p(x_1\dots x_{12})}$$
$$= \frac{\sum_{x_2}\psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)\psi(x_4)m(x_4)}{\sum_{x_1}\sum_{x_2}\psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)\psi(x_4)m(x_4)}$$
$$= \frac{\sum_{x_2}\psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)}{\sum_{x_1}\sum_{x_2}\psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4)\psi(x_1)} = \frac{\sum_{x_2}\psi(x_1=1,x_2)\psi(x_2)\psi(x_2,x_4=1)\psi(x_1=1)}{\sum_{x_1}\sum_{x_2}\psi(x_1,x_2)\psi(x_2)\psi(x_2,x_4=1)\psi(x_1)}$$
$$= \frac{5\times 10\times 5\times 2 + 20\times 2\times 20\times 2}{20\times 10\times 5\times 10 + 5\times 2\times 20\times 10 + 5\times 10\times 5\times 2 + 20\times 2\times 20\times 2} = \frac{7}{47} = 0.1489$$

Solution: **Approach 2:** Node x_1 is conditionally independent of all other nodes except node x_2, given x_4. Let's compute the message from x_4 to x_2.

$m_{x_4\to x_2}$		
$x_2 = 0$	$0\times 20 + 2\times 5$	10
$x_2 = 1$	$0\times 5 + 2\times 20$	40

Now, let's compute the message from x_2 to x_1.

$m_{x_2\to x_1}$		
$x_1 = 0$	$10\times 20\times 10 + 2\times 5\times 40$	2400
$x_1 = 1$	$10\times 5\times 10 + 2\times 20\times 40$	2100

$$\tilde{p}(x_1=\hat{x}_1|x_3=0,x_4=1,x_6=0) = \psi(x_1=\hat{x}_1)\times m_{x_2\to x_1}(x_1=\hat{x}_1) \tag{5}$$

$\tilde{p}$		
$x_1 = 0$	10×2400	24000
$x_1 = 1$	2×2100	4200

$$p(x_1=1|x_3=0,x_4=1,x_6=0) = \frac{4200}{24000+4200} = \frac{7}{47} = 0.1489 \tag{6}$$

Problem 1.c. [2 points] Compute $p(x_{10}=1|x_9=1,x_{12}=1,x_2=0)$.

Solution: **Approach 1:** According to the conditional independence in MRF.

$$p(x_{10}|x_9,x_{12},x_2) = p(x_{10}|x_9,x_{12}) = \frac{p(x_{10},x_9,x_{12})}{p(x_9,x_{12})}$$

Denote message from nodes $M_{E/x_9,x_{10},x_{11},x_{12}} = \{x_1,\dots x_8\}$ to x_9 as

$$m(x_9) = \sum_{i,j\in M_{E/x_9,x_{10},x_{11},x_{12}}} \psi(x_i)\psi(x_i,x_j)$$

Then

$$p(x_{10}|x_9,x_{12},x_2) = \frac{p(x_{10},x_9,x_{12})}{p(x_9,x_{12})} = \frac{\sum_{x_{11}}\sum_{i,j\in M_{E/x_9,x_{10},x_{11},x_{12}}} p(x_1\dots x_{12})}{\sum_{x_{10}}\sum_{x_{11}}\sum_{i,j\in M_{E/x_9,x_{10},x_{11},x_{12}}} p(x_1\dots x_{12})}$$

$$= \frac{\sum_{x_{11}} \psi(x_9)\psi(x_9,x_{10})\psi(x_{10})\psi(x_{10},x_{11})\psi(x_{11})\psi(x_{10},x_{12})\psi(x_{12})m(x_9)}{\sum_{x_{10}}\sum_{x_{11}} \psi(x_9)\psi(x_9,x_{10})\psi(x_{10})\psi(x_{10},x_{11})\psi(x_{11})\psi(x_{10},x_{12})\psi(x_{12})m(x_9)}$$

$$= \frac{\sum_{x_{11}} \psi(x_9)\psi(x_9,x_{10})\psi(x_{10})\psi(x_{10},x_{11})\psi(x_{11})\psi(x_{10},x_{12})\psi(x_{12})m(x_9)}{\sum_{x_{10}}\sum_{x_{11}} \psi(x_9)\psi(x_9,x_{10})\psi(x_{10})\psi(x_{10},x_{11})\psi(x_{11})\psi(x_{10},x_{12})\psi(x_{12})m(x_9)}$$

$$= \frac{\sum_{x_{11}} \psi(x_9=1,x_{10}=1)\psi(x_{10}=1)\psi(x_{10}=1,x_{11})\psi(x_{11})\psi(x_{10}=1,x_{12}=1)}{\sum_{x_{10}}\sum_{x_{11}} \psi(x_9=1,x_{10})\psi(x_{10})\psi(x_{10},x_{11})\psi(x_{11})\psi(x_{10},x_{12}=1)}$$

$$= \frac{20\times 2\times 5\times 10\times 20 + 20\times 2\times 20\times 2\times 20}{5\times 10\times 20\times 10\times 5 + 20\times 2\times 5\times 10\times 20 + 5\times 10\times 5\times 2\times 5 + 20\times 2\times 20\times 2\times 20}$$

$$= \frac{48}{83} = 0.5783$$

Solution: Approach 2:

Node x_{10} is conditionally independent of all other nodes except node x_{11}, given x_9 and x_{12}. Let's compute messages from x_{11}, x_9, and x_{12} to x_{10}.

$m_{x_{11}\to x_{10}}$	
$x_{10}=0$	$10\times 20 + 2\times 5 = 210$
$x_{10}=1$	$10\times 5 + 2\times 20 = 90$

$m_{x_9\to x_{10}}$	
$x_{10}=0$	$0\times 20 + 2\times 5 = 10$
$x_{10}=1$	$0\times 5 + 2\times 20 = 40$

$m_{x_{12}\to x_{10}}$	
$x_{10}=0$	$0\times 20 + 2\times 5 = 10$
$x_{10}=1$	$0\times 5 + 2\times 20 = 40$

$$\tilde{p}(x_{10}=\hat{x}_{10}|x_9=1,x_{12}=1,x_2=0) = \psi(x_{10}=\hat{x}_{10})\times m_{x_{11}\to x_{10}}(x_{10}=\hat{x}_{10}) \quad (7)$$
$$\times\, m_{x_9\to x_{10}}(x_{10}=\hat{x}_{10})\times m_{x_{12}\to x_{10}}(x_{10}=\hat{x}_{10}) \quad (8)$$

$\tilde{p}$		
$x_{10}=0$	$10\times 210\times 10\times 10$	210000
$x_{10}=1$	$2\times 90\times 40\times 40$	288000

$$p(x_{10}=1|x_9=1,x_{12}=1,x_2=0) = \frac{288000}{210000+288000} = \frac{48}{83} = 0.5783 \quad (9)$$

Problem 1.d. [2 points] Compute $p(x_6=0|x_4=1,x_8=1,x_{10}=0)$.

Solution: Approach 1: According to the conditional independence in MRF.

$$p(x_6|x_4,x_8,x_{10}) = p(x_6|x_4,x_8) = \frac{\sum_{x_7} p(x_4,x_6,x_8,x_7)}{\sum_{x_7}\sum_{x_7} p(x_4,x_6,x_8,x_7)}$$

Similar to previous questions, we can ignore the message from outer nodes to x_4 and x_8 (since the messages to this two nodes appear in both denominator and numerator, so they can be eliminated). Then

$$p(x_6 = 0|x_4 = 1, x_8 = 1) = \frac{\sum_{x_7} \psi(x_4 = 1, x_6 = 0)\psi(x_8 = 1, x_6 = 0)\psi(x_7)\psi(x_7, x_6 = 0)\psi(x_6 = 0)}{\sum_{x_6}\sum_{x_7} \psi(x_4 = 1, x_6)\psi(x_8 = 1, x_6)\psi(x_7)\psi(x_7, x_6)\psi(x_6)}$$

$$= \frac{10 \times 5 \times 5 \times 2 \times 5 + 5 \times 5 \times 10 \times 20 \times 10}{10 \times 5 \times 5 \times 2 \times 5 + 5 \times 5 \times 10 \times 20 \times 10 + 20 \times 20 \times 10 \times 5 \times 2 + 20 \times 20 \times 2 \times 20 \times 2} = \frac{35}{83} = 0.4217$$

Solution: **Approach 2:** Node x_6 is conditionally independent of all other nodes except node x_7, given x_4 and x_8. Let's compute messages from x_7, x_4, and x_8 to x_6.

$m_{x_7 \to x_6}$	
$x_6 = 0$	$10 \times 20 + 2 \times 5 = 210$
$x_6 = 1$	$10 \times 5 + 2 \times 20 = 90$

$m_{x_4 \to x_6}$	
$x_6 = 0$	$0 \times 20 + 2 \times 5 = 10$
$x_6 = 1$	$0 \times 5 + 2 \times 20 = 40$

$m_{x_8 \to x_6}$	
$x_6 = 0$	$0 \times 20 + 2 \times 5 = 10$
$x_6 = 1$	$0 \times 5 + 2 \times 20 = 40$

$$\tilde{p}(x_6 = \hat{x}_6|x_4 = 1, x_8 = 1, x_{10} = 0) = \psi(x_6 = \hat{x}_6) \times m_{x_7 \to x_6}(x_6 = \hat{x}_6) \tag{10}$$

$$\times\, m_{x_8 \to x_6}(x_6 = \hat{x}_6) \times m_{x_8 \to x_6}(x_6 = \hat{x}_6) \tag{11}$$

$\tilde{p}$		
$x_6 = 0$	$10 \times 210 \times 10 \times 10$	210000
$x_6 = 1$	$2 \times 90 \times 40 \times 40$	288000

$$p(x_6 = 0|x_4 = 1, x_8 = 1, x_{10} = 0) = \frac{210000}{210000 + 288000} = \frac{35}{83} = 0.4217 \tag{12}$$

Problem 1.e. [2 points] Compute $p(x_8 = 1|x_1 = 0, x_6 = 0, x_9 = 1, x_{12} = 1)$.

Solution: **Approach 1:** According to the conditional independence in MRF.

$$p(x_8|x_1, x_6, x_9, x_{12}) = p(x_8|x_6, x_9) = \frac{p(x_8, x_6, x_9)}{\sum_{x_8} p(x_8, x_6, x_9)}$$

Similar to previous questions, we can ignore the message from outer nodes to x_6 and x_9 (since the messages to this two nodes appear in both denominator and numerator, so they can be eliminated). Then

$$p(x_8 = 1|x_6 = 0, x_9 = 1) = \frac{\psi(x_8 = 1, x_6 = 0)\psi(x_9 = 1, x_8 = 1)\psi(x_8 = 1)}{\sum_{x_8} \psi(x_8, x_6 = 0)\psi(x_9 = 1, x_8)\psi(x_8)}$$

$$= \frac{5 \times 20 \times 2}{5 \times 20 \times 2 + 20 \times 5 \times 10} = \frac{1}{6} = 0.1667$$

Approach 2: Node x_8 is conditionally independent of all other nodes, given x_6 and x_9. Let's compute messages from x_6 and x_9 to x_8.

$m_{x_6 \to x_8}$		
$x_8 = 0$	$10 \times 20 + 0 \times 5$	200
$x_8 = 1$	$10 \times 5 + 0 \times 20$	50

$m_{x_9 \to x_8}$		
$x_8 = 0$	$0 \times 20 + 2 \times 5$	10
$x_8 = 1$	$0 \times 5 + 2 \times 20$	40

$$\tilde{p}(x_8 = \hat{x}_8 | x_1 = 0, x_6 = 0, x_9 = 1, x_{12} = 1) = \psi(x_8 = \hat{x}_8) \times m_{x_6 \to x_8}(x_8 = \hat{x}_8) \tag{13}$$
$$\times\, m_{x_9 \to x_8}(x_8 = \hat{x}_8) \tag{14}$$

$\tilde{p}$		
$x_8 = 0$	$10 \times 200 \times 10$	20000
$x_8 = 1$	$2 \times 50 \times 40$	4000

$$p(x_8 = 1 | x_1 = 0, x_6 = 0, x_9 = 1, x_{12} = 1) = \frac{4000}{20000 + 4000} = \frac{1}{6} = 0.1667 \tag{15}$$

Problem 1.f. [2 points] Compute $p(x_2 = 0 | x_1 = 0, x_3 = 1, x_4 = 1, x_7 = 1, x_{11} = 0)$.

Solution: **Approach 1:** According to the conditional independence in MRF.

$$p(x_2 | x_1, x_3, x_4, x_7, x_{11}) = p(x_2 | x_1, x_4) = \frac{p(x_1, x_2, x_4)}{\sum_{x_2} p(x_1, x_2, x_4)}$$

Similar to previous questions, we can ignore the message from outer nodes to x_1 and x_4 (since the messages to this two nodes appear in both denominator and numerator, so they can be eliminated). Then

$$p(x_2 | x_1, x_4) = \frac{\psi(x_2 = 0, x_1 = 0)\psi(x_4 = 1, x_2 = 0)\psi(x_2 = 0)}{\sum_{x_2} \psi(x_2, x_1 = 0)\psi(x_4 = 1, x_2)\psi(x_2)}$$
$$= \frac{20 \times 5 \times 10}{5 \times 20 \times 2 + 20 \times 5 \times 10} = \frac{5}{6} = 0.8333$$

Solution: **Approach 2:**

Node x_2 is conditionally independent of all other nodes, given x_1 and x_4. This will result in a similar structure and, therefore, exactly the same computations as the previous question. However, here we need to compute the probability for $x_2 = 0$ whereas in the previous question it was for $x_8 = 1$. The answer can then be computed as $1 - \frac{1}{6} = \frac{5}{6} = 0.8333$.

Problem 2. (Linear Gaussian)

For this tutorial problem, we will consider a specific DGM that is the basis for more sophisticated models such as Probabilistic PCA and Linear Dynamical Systems. This model is called the Linear-Gaussian Model. *Note:* for this problem, we will be denoting random variables with lower case letters, and bolded lowercase letters to represent vectors, and bolded uppercase letters to represent matrices.

Problem 2.a. We will build our way up towards this model. As a prelude, consider K *independent* univariate Gaussian random variables $x_1, x_2, \ldots, x_K$,

$$p(x_k) = \mathcal{N}(\mu_k, \sigma_k^2)$$

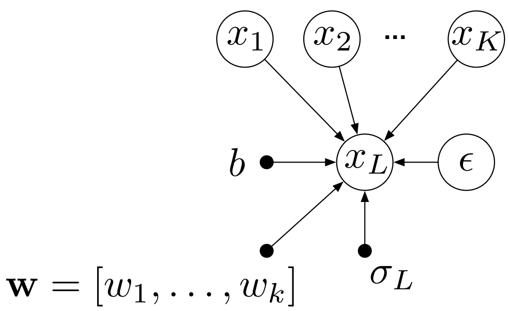
for $k = 1, 2, \ldots, K$. Define the random variable x_L,

$$x_L = b + \sigma_L \epsilon + \sum_{k=1}^{K} w_k x_k$$

where $\epsilon \sim \mathcal{N}(0, 1)$.

1. Draw out the DGM for the model described above.
2. Show that $p(x_L|x_1, \ldots, x_K) = \mathcal{N}\left(b + \sum_{k=1}^{K} w_k x_k, \sigma_L^2\right)$. In other words, x_L is Gaussian distributed with mean $b + \sum_{k=1}^{K} w_k x_k$ and variance σ_L^2.
3. Define the random variable $\mathbf{x} = (x_1, x_2, \ldots, x_K, x_L)$. Show that $\mathbf{x}$ is a *multivariate* Gaussian random variable. *Hint: Consider the definition of the multivariate Gaussian and the properties of Gaussians.*

Solution:



Solution: Since ϵ follows Gaussian distribution, then the linear transformation of ϵ plus a constant is still Gaussian. As such, we just need to compute mean and variance of the Gaussian distribution.

$$\mathbb{E}\left[x_L|x_1, \cdots, x_K\right] = \mathbb{E}\left[b + \sigma_L \epsilon + \sum_{k=1}^{K} w_k x_k\right] \tag{16}$$

$$= \mathbb{E}[b] + \sigma_L \mathbb{E}[\epsilon] + \sum_{k=1}^{K} w_k \mathbb{E}[x_k] \tag{17}$$

Since $x_1, \cdots, x_K$ are known, $\mathbb{E}[x_k] = x_k$, therefore $\sum_{k=1}^{K} w_k \mathbb{E}[x_k] = \sum_{k=1}^{K} w_k x_k$. Since b is a constant, $\mathbb{E}[b] = b$. Also, given $\mathbb{E}[\epsilon] = 0$. Thus,

$$\mathbb{E}\left[x_L|x_1, \cdots, x_K\right] = b + \sum_{k=1}^{K} w_k x_k \tag{18}$$

Similarly, for variance, we have

$$\begin{aligned}
\mathrm{Var}\left[x_L|x_1,\cdots,x_K\right] &= \mathrm{Var}\left[b+\sigma_L\epsilon+\sum_{k=1}^{K}w_kx_k\right] &(19)\\
&= \mathrm{Var}[b]+\sigma_L^2\mathrm{Var}[\epsilon]+\sum_{k=1}^{K}w_k^2\mathrm{Var}[x_k] &(20)\\
&= \sigma_L^2\mathrm{Var}[\epsilon]=\sigma_L^2 &(21)
\end{aligned}$$

Solution: Since

$$p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k\in\mathbf{x}_{\pi_i}}\frac{1}{\sigma_k^2}(x_k-\mu_k)^2\right]\right\} \tag{22}$$

$$p(x_L|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_L^2}(x_L-(b+\sum_{k\in\mathbf{x}_{\pi_i}}w_kx_k))^2\right]\right\} \tag{23}$$

$$p(\mathbf{x}_{\pi_i})p(x_L|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k\in\mathbf{x}_{\pi_i}}\frac{1}{\sigma_k^2}(x_k-\mu_k)^2+\frac{1}{\sigma_L^2}(x_L-(b+\sum_{k\in\mathbf{x}_{\pi_i}}w_kx_k))^2\right]\right\} \tag{24}$$

Denote $\mathbf{x}_{\pi_i}=(x_1,\cdots,x_K)$, $\mathbf{w}=[w_1,\cdots,w_K]^T$.

We look at the terms in exponential,

$$\begin{aligned}
&\sum_{k\in\mathbf{x}_{\pi_i}}\frac{1}{\sigma_k^2}(x_k-\mu_k)^2+\frac{1}{\sigma_L^2}(x_L-(b+\sum_{k\in\mathbf{x}_{\pi_i}}w_kx_k))^2 &(25)\\
=&(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\boldsymbol{\Sigma}_{\pi_i}^{-1}(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})+\left(x_L-(b+\mathbf{w}^T\mathbf{x}_{\pi_i})\right)^T\boldsymbol{\Sigma}_L^{-1}\left(x_L-(b+\mathbf{w}^T\mathbf{x}_{\pi_i})\right) &(26)\\
=&(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\boldsymbol{\Sigma}_{\pi_i}^{-1}(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}) &(27)\\
&+\left(x_L-(b+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})-\mathbf{w}^T(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})\right)^T\boldsymbol{\Sigma}_L^{-1}\left(x_L-(b+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})-\mathbf{w}^T(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})\right) &(28)\\
=&(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\left[\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T\right](\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}) &(29)\\
&+\left(x_L-(b+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})\right)^T\boldsymbol{\Sigma}_L^{-1}\left(x_L-(b+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})\right)-2(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}(x_L-\mu_L) &(30)\\
=&\begin{bmatrix}\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}\\ x_L-\mu_L\end{bmatrix}^T\begin{bmatrix}\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T\\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}\\ x_L-\mu_L\end{bmatrix} &(31)
\end{aligned}$$

where $\mu_L=b+\mathbf{w}^T\mathbf{x}_{\pi_i}$, $\boldsymbol{\Sigma}_L=[\sigma_L^2]$ and $\boldsymbol{\Sigma}_{\pi_i}=\mathrm{diagonal}(\sigma_1^2,\cdots,\sigma_K^2)$.

Now we verify

$$\begin{bmatrix}\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T\\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1}\end{bmatrix}$$

is positive definite. For any $\mathbf{u}$, we have

$$\begin{aligned}
\mathbf{u}^T\begin{bmatrix}\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T\\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1}\end{bmatrix}\mathbf{u} =&\mathbf{u}^T\begin{bmatrix}1 & 0\\ -\boldsymbol{w}^T & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_L^{-1} & 0\\ 0 & \boldsymbol{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\boldsymbol{w}\\ 0 & 1\end{bmatrix}\mathbf{u} &(32)\\
=&\left[\mathbf{u}\begin{bmatrix}1 & -\boldsymbol{w}\\ 0 & 1\end{bmatrix}\right]^T\begin{bmatrix}\boldsymbol{\Sigma}_L^{-1} & 0\\ 0 & \boldsymbol{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\left[\begin{bmatrix}1 & -\boldsymbol{w}\\ 0 & 1\end{bmatrix}\mathbf{u}\right]>0 &(33)
\end{aligned}$$

The last inequality is given by that $\mathbf{\Sigma}_L$ and $\mathbf{\Sigma}_{\pi_i}$ are positive definite.

Therefore,

$$\begin{bmatrix} \mathbf{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\mathbf{\Sigma}_L^{-1}\boldsymbol{w}^T & \mathbf{\Sigma}_L^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\mathbf{\Sigma}_L^{-1} & \mathbf{\Sigma}_L^{-1} \end{bmatrix}$$

is a valid covariance matrix.

Define

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\pi_i} \\ b + \mathbf{w}^T\mathbf{x}_{\pi_i} \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\mathbf{\Sigma}_L^{-1}\boldsymbol{w}^T & \mathbf{\Sigma}_L^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\mathbf{\Sigma}_L^{-1} & \mathbf{\Sigma}_L^{-1} \end{bmatrix}^{-1}$$

We have

$$p(x_L, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_{\pi_i})p(x_L|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left(\begin{bmatrix}\mathbf{x}_{\pi_i} \\ x_L\end{bmatrix} - \boldsymbol{\mu}\right)^T \mathbf{\Sigma}^{-1}\left(\begin{bmatrix}\mathbf{x}_{\pi_i} \\ x_L\end{bmatrix} - \boldsymbol{\mu}\right)\right\} \tag{34}$$

Therefore, $p(x_L, \mathbf{x}_{\pi_i})$ is Gaussian distributed.

Problem 2.b. Let's now move to the more complex case. Consider an *arbitrary* DGM G where each node j without any parents is Gaussian distributed with mean μ_j and variance σ_j^2. The remaining nodes are defined as

$$x_i = b_i + \left(\sum_{j \in x_{\pi_i}} w_{i,j}x_j\right) + \sigma_i\epsilon_i$$

where x_{π_i} denotes the set of node i's parents and ϵ_i is the standard normal random variable $\epsilon_i \sim \mathcal{N}(0, 1)$.

1. Show that each node x_i has the conditional distribution: $p(x_i|x_{\pi_i}) = \mathcal{N}\left(b_i + \sum_{j \in x_{\pi_i}} w_{i,j}x_j, \sigma_i^2\right)$
2. Define the random variable $\mathbf{x} = (x_1, x_2, \ldots, x_D)$. Show that $\mathbf{x}$ is a *multivariate Gaussian.*

Solution: We just apply the same derivation from problem 2.a.

Since ϵ_i follows Gaussian distribution, then the linear transformation of ϵ_i plus constant is still Gaussian distribution. Then we just need to compute mean and variance of the Gaussian distribution.

$$\begin{aligned} \mathbb{E}\left[x_i|x_{\pi_i}\right] &= \mathbb{E}\left[b_i + \sigma_i\epsilon_i + \sum_{j \in x_{\pi_i}} w_jx_j\right] && (35) \\ &= \mathbb{E}[b_i] + \sigma_i\mathbb{E}[\epsilon_i] + \sum_{j \in x_{\pi_i}} w_j\mathbb{E}[x_j] && (36) \\ & && (37) \end{aligned}$$

Since x_{π_i} are known, $\mathbb{E}[x_j] = x_j, (j \in x_{\pi_i})$, therefore $\sum_{j \in x_{\pi_i}} w_j\mathbb{E}[x_j] = \sum_{k \in x_{\pi_i}} w_jx_j$. Since b_i is a constant, $\mathbb{E}[b_i] = b_i$. Also, given $\mathbb{E}[\epsilon_i] = 0$. Thus,

$$\mathbb{E}\left[x_i|x_{\pi_i}\right] = b_i + \sum_{j \in x_{\pi_i}} w_jx_j \tag{38}$$

Similarly, for variance, we have

$$\begin{aligned}
\mathrm{Var}\left[x_i|x_{\pi_i}\right] &= \mathrm{Var}\left[b_i + \sigma_i\epsilon_i + \sum_{k\in x_{\pi_i}} w_j x_j\right] && (39)\\
&= \mathrm{Var}[b_i] + \sigma_i^2\mathrm{Var}[\epsilon_i] + \sum_{k\in x_{\pi_i}} w_j^2\mathrm{Var}[x_j] && (40)\\
&= \sigma_i^2\mathrm{Var}[\epsilon_i] = \sigma_i^2 && (41)
\end{aligned}$$

Solution: We know that $p(x_i|\mathbf{x}_{\pi_i})$ is Gaussian distributed. Let's first assume $p(\mathbf{x}_{\pi_i})$ is also Gaussian. Next, we prove $p(x_i, \mathbf{x}_{\pi_i}) = p(x_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i})$ is also Gaussian. We can perform induction from root node following the topological order in the graph to show that the joint distribution of all nodes is a multivariate Gaussian.

Since

$$p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k\in\mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k-\mu_k)^2\right]\right\} \tag{42}$$

$$p(x_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_i^2}(x_i-(b_i+\sum_{k\in\mathbf{x}_{\pi_i}} w_k x_k))^2\right]\right\} \tag{43}$$

$$p(\mathbf{x}_{\pi_i})p(x_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k\in\mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k-\mu_k)^2 + \frac{1}{\sigma_i^2}(x_i-(b_i+\sum_{k\in\mathbf{x}_{\pi_i}} w_k x_k))^2\right]\right\} \tag{44}$$

We look at the terms in exponential and denote $\mathbf{w} = [w_1, \cdots, w_{K_{\pi_i}}]^T$

$$\begin{aligned}
&\sum_{k\in\mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k-\mu_k)^2 + \frac{1}{\sigma_i^2}(x_i-(b_i+\sum_{k\in\mathbf{x}_{\pi_i}} w_k x_k))^2 && (45)\\
=&(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\boldsymbol{\Sigma}_{\pi_i}^{-1}(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}) + \left(x_i-(b_i+\mathbf{w}^T\mathbf{x}_{\pi_i})\right)^T\boldsymbol{\Sigma}_i^{-1}\left(x_i-(b_i+\mathbf{w}^T\mathbf{x}_{\pi_i})\right) && (46)\\
=&(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\boldsymbol{\Sigma}_{\pi_i}^{-1}(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}) && (47)\\
&+\left(x_i-(b_i+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})-\mathbf{w}^T(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})\right)^T\boldsymbol{\Sigma}_i^{-1}\left(x_i-(b_i+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})-\mathbf{w}^T(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})\right) && (48)\\
=&(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\left[\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T\right](\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}) && (49)\\
&+\left(x_i-(b_i+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})\right)^T\boldsymbol{\Sigma}_i^{-1}\left(x_i-(b_i+\mathbf{w}^T\boldsymbol{\mu}_{\pi_i})\right) - 2(\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i})^T\boldsymbol{w}\boldsymbol{\Sigma}_i^{-1}(x_i-\mu_i) && (50)\\
=&\begin{bmatrix}\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}\\ x_i-\mu_i\end{bmatrix}^T\begin{bmatrix}\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T\\ \boldsymbol{w}\boldsymbol{\Sigma}_i^{-1} & \boldsymbol{\Sigma}_i^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{\pi_i}-\boldsymbol{\mu}_{\pi_i}\\ x_i-\mu_i\end{bmatrix} && (51)
\end{aligned}$$

where $\mu_i = b_i + \mathbf{w}^T\mathbf{x}_{\pi_i}$, $\boldsymbol{\Sigma}_i = \sigma_i^2$

Now we verify

$$\begin{bmatrix}\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T\\ \boldsymbol{w}\boldsymbol{\Sigma}_i^{-1} & \boldsymbol{\Sigma}_i^{-1}\end{bmatrix}$$

is positive definite. For any $\mathbf{u}$, we have

$$\begin{aligned}
\mathbf{u}^T\begin{bmatrix}\boldsymbol{\Sigma}_{\pi_i}^{-1}+\boldsymbol{w}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_i^{-1}\boldsymbol{w}^T\\ \boldsymbol{w}\boldsymbol{\Sigma}_i^{-1} & \boldsymbol{\Sigma}_i^{-1}\end{bmatrix}\mathbf{u} =& \mathbf{u}^T\begin{bmatrix}1 & 0\\ -\boldsymbol{w}^T & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_i^{-1} & 0\\ 0 & \boldsymbol{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\boldsymbol{w}\\ 0 & 1\end{bmatrix}\mathbf{u} && (52)\\
=&\left[\mathbf{u}\begin{bmatrix}1 & -\boldsymbol{w}\\ 0 & 1\end{bmatrix}\right]^T\begin{bmatrix}\boldsymbol{\Sigma}_i^{-1} & 0\\ 0 & \boldsymbol{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\left[\begin{bmatrix}1 & -\boldsymbol{w}\\ 0 & 1\end{bmatrix}\mathbf{u}\right] > 0 && (53)
\end{aligned}$$

The last inequality is given by that $\mathbf{\Sigma}_i$ and $\mathbf{\Sigma}_{\pi_i}$ are positive definite.

Therefore,

$$\begin{bmatrix} \mathbf{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\mathbf{\Sigma}_i^{-1}\boldsymbol{w}^T & \mathbf{\Sigma}_i^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\mathbf{\Sigma}_i^{-1} & \mathbf{\Sigma}_i^{-1} \end{bmatrix}$$

is a valid covariance matrix.

Define

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\pi_i} \\ b_i + \mathbf{w}^T\mathbf{x}_{\pi_i} \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\mathbf{\Sigma}_i^{-1}\boldsymbol{w}^T & \mathbf{\Sigma}_i^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\mathbf{\Sigma}_i^{-1} & \mathbf{\Sigma}_i^{-1} \end{bmatrix}^{-1}$$

We have

$$p(x_i, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_{\pi_i})p(x_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{ -\frac{1}{2}\left(\begin{bmatrix}\mathbf{x}_{\pi_i} \\ x_i\end{bmatrix} - \boldsymbol{\mu}\right)^T \mathbf{\Sigma}^{-1}\left(\begin{bmatrix}\mathbf{x}_{\pi_i} \\ x_i\end{bmatrix} - \boldsymbol{\mu}\right)\right\} \tag{54}$$

Therefore, $p(x_i, \mathbf{x}_{\pi_i})$ is Gaussian distributed. We start from the nodes without parents and analyze each node in topological order, and by induction, we can show that the joint distribution of all nodes is multivariate Gaussian.

Problem 2.c. We can determine the mean of $\mathbf{x}$ using a recursive method. Note that $\mathbb{E}[\mathbf{x}] = (\mathbb{E}[x_1], \ldots, \mathbb{E}[x_D])^\top$. Show that the expectation of each component $\mathbb{E}[x_i]$ is given by:

$$\mathbb{E}[x_i] = b_i + \sum_{j \in x_{\pi_i}} w_{i,j}\mathbb{E}[x_j]$$

Solution: Since $x_i = b_i + \left(\sum_{j\in x_{\pi_i}} w_{i,j}x_j\right) + \sigma_i\epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0,1)$

$$\begin{aligned} \mathbb{E}[x_i] &= \mathbb{E}\left[b_i + \left(\sum_{j\in x_{\pi_i}} w_{i,j}x_j\right) + \sigma_i\epsilon_i\right] && (55) \\ &= \mathbb{E}[b_i] + \sum_{j\in x_{\pi_i}} w_{i,j}\mathbb{E}[(x_j)] + \sigma_i\mathbb{E}[\epsilon_i] && (56) \\ &= b_i + \sum_{j\in x_{\pi_i}} w_{i,j}\mathbb{E}[(x_j)] && (57) \end{aligned}$$

Problem 2.d. Likewise, we can determine the covariance matrix of $\mathbf{x}$. Note that

$$\mathbf{\Sigma}_{ij} = \mathrm{Cov}[x_i, x_j] = \mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$$

1. Show that $\mathrm{Cov}[x_i, x_j] = I_{ij}\sigma_j^2 + \sum_{k\in x_{\pi_j}} w_{j,k}\mathrm{Cov}[x_i, x_k]$
2. If the DGM G has no edges, is the covariance matrix $\mathbf{\Sigma}$ a spherical, diagonal, or general symmetric covariance matrix? How many parameters does it have?

3. If the DGM G is fully-connected, what kind of matrix is the covariance matrix $\mathbf{\Sigma}$? Is it spherical, diagonal, or a general symmetric covariance matrix? How many parameters does it have?

Solution: Since $x_j = b_j + \left(\sum_{k \in x_{\pi_j}} w_{j,k} x_k\right) + \sigma_j \epsilon_j$

$$\begin{aligned}
\mathrm{Cov}[x_i, x_j] &= \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,[x_j - \mathbb{E}[x_j]]\right] && (58)\\
&= \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\left[b_j + \left(\sum_{k \in x_{\pi_j}} w_{j,k} x_k\right) + \sigma_j \epsilon_j - \mathbb{E}[x_j]\right]\right] && (59)\\
&= \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\left[b_j + \left(\sum_{k \in x_{\pi_j}} w_{j,k} x_k\right) + \sigma_j \epsilon_j - \left(b_j + \sum_{k \in x_{\pi_j}} w_{j,k} \mathbb{E}[x_k]\right)\right]\right] && (60)\\
&= \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\left[\sum_{k \in x_{\pi_j}} w_{j,k} \left(x_k - \mathbb{E}[x_k]\right) + \sigma_j \epsilon_j\right]\right] && (61)\\
&= \sum_{k \in x_{\pi_j}} w_{j,k} \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,[x_k - \mathbb{E}[x_k]]\right] + \sigma_j \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,\epsilon_j\right] && (62)\\
&= \sum_{k \in x_{\pi_j}} w_{j,k} \mathrm{Cov}\left(x_i, x_k\right) + \sigma_j \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,\epsilon_j\right] && (63)
\end{aligned}$$

Since $x_i = b_i + \left(\sum_{j \in x_{\pi_i}} w_{i,j} x_j\right) + \sigma_i \epsilon_i$

$$\begin{aligned}
\sigma_j \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,\epsilon_i\right] &= \sigma_j \mathbb{E}\left[\left[\sum_{k \in x_{\pi_i}} w_{i,k} \left(x_k - \mathbb{E}[x_k]\right) + \sigma_i \epsilon_i\right] \epsilon_j\right] && (64)\\
& && (65)
\end{aligned}$$

Since x_k and ϵ_j are independent, therefore,

$$\begin{aligned}
\mathbb{E}\left[\left[\sum_{k \in x_{\pi_i}} w_{i,k} \left(x_k - \mathbb{E}[x_k]\right)\right] \epsilon_j\right] &= \mathbb{E}\left[\sum_{k \in x_{\pi_i}} w_{i,k} \left(x_k - \mathbb{E}[x_k]\right)\right] \mathbb{E}\left[\epsilon_j\right] && (66)\\
&= \mathbb{E}\left[\sum_{k \in x_{\pi_i}} w_{i,k} \left(x_k - \mathbb{E}[x_k]\right)\right] \cdot 0 && (67)\\
&= 0 && (68)
\end{aligned}$$

Then, if $i = j$ we have

$$\begin{aligned}
\sigma_i \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,\epsilon_i\right] &= \sigma_i \sigma_i \mathbb{E}\left[\epsilon_i \epsilon_i\right] && (69)\\
&= \sigma_i^2 \mathbb{E}\left[\epsilon_i \epsilon_i\right] && (70)\\
&= \sigma_i^2 \mathbb{E}\left[[\epsilon_i - \mathbb{E}[\epsilon_i]]\,[\epsilon_i - \mathbb{E}[\epsilon_i]]\right] && (71)\\
&= \sigma_i^2 \mathrm{Var}[\epsilon_i] = \sigma_i^2 && (72)\\
& && (73)
\end{aligned}$$

If $i \neq j$, we have ϵ_i and ϵ_j independent.

$$\begin{aligned}
\sigma_i \mathbb{E}\left[[x_i - \mathbb{E}[x_i]]\,\epsilon_i\right] &= \sigma_i \sigma_j \mathbb{E}\left[\epsilon_i \epsilon_j\right] && (74)\\
&= \sigma_i \sigma_j \mathbb{E}\left[\epsilon_i\right] \mathbb{E}\left[\epsilon_j\right] && (75)\\
&= 0 && (76)\\
& && (77)
\end{aligned}$$

Put all these together,

$$\mathrm{Cov}[x_i, x_j] = I_{ij}\sigma_j^2 + \sum_{k \in x_{\pi_j}} w_{j,k} \mathrm{Cov}[x_i, x_k] \tag{78}$$

where I_{ij} equals 1, if $i = j$; and equals 0, if $i \neq j$.

Solution: If G has no edges, the covariance matrix is a diagonal matrix and the number of parameters is D.

Solution: If G is fully-connected, the covariance matrix is a general symmetric matrix. The number of parameters is $\frac{D(D-1)}{2}$.

Problem 2.e. (Challenge) Consider now the situation where each node in G is a multivariate Gaussian random variable. More concretely, each node j without parents is multivariate Gaussian distributed with mean $\boldsymbol{\mu}_j$ and variance $\boldsymbol{\Sigma}_j$. The conditional for the remaining nodes are also multivariate Gaussian:

$$p(\mathbf{x}_i|\mathbf{x}_{\pi_i}) = \mathcal{N}\left(\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j, \boldsymbol{\Sigma}_i\right)$$

Show that the joint distribution over *all* variables is multivariate Gaussian.

Solution: First assume $\mathbf{x}_{\pi_i} \sim \mathcal{N}(\mu_{\pi i}, \boldsymbol{\Sigma}_{\pi_i})$, and we show that $p(\mathbf{x}_i, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i})$ is Gaussian distributed. Then, we apply this property on the graph G. We start from root nodes and, by induction, we can prove the whole joint distribution is a multivariate Gaussian.

Since

$$p(\mathbf{x}_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)\right\} \tag{79}$$

$$p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \boldsymbol{\Sigma}_{\pi_i}^{-1}\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)\right\} \tag{80}$$

$$\begin{aligned}
p(\mathbf{x}_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i}) \propto \exp\{&-\frac{1}{2}\left(\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \boldsymbol{\Sigma}_{\pi_i}^{-1}\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)\right) && (81)\\
&-\frac{1}{2}\left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)\} && (82)\\
& && (83)
\end{aligned}$$

We analyze the terms in the exponential.

$$\left(\mathbf{x}_{\pi_i}-\mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1}\left(\mathbf{x}_{\pi_i}-\mu_{\pi i}\right)+\left(\mathbf{x}_i-(\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\mathbf{x}_j)\right)^T \mathbf{\Sigma}_i^{-1}\left(\mathbf{x}_i-(\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\mathbf{x}_j)\right)\} \tag{84}$$

$$=\left(\mathbf{x}_{\pi_i}-\mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1}\left(\mathbf{x}_{\pi_i}-\mu_{\pi i}\right) \tag{85}$$

$$+\left(\mathbf{x}_i-\left((\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j)+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}(\mathbf{x}_j-\boldsymbol{\mu}_j)\right)\right)^T \tag{86}$$

$$\mathbf{\Sigma}_i^{-1}\left(\mathbf{x}_i-\left((\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j)+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}(\mathbf{x}_j-\boldsymbol{\mu}_j)\right)\right) \tag{87}$$

$$=\left[\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix}-\begin{bmatrix}\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}\right]^T\begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix} \tag{88}$$

$$\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\left[\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix}-\begin{bmatrix}\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}\right] \tag{89}$$

where $\mathbf{W_i}=[\mathbf{W}_{i,1},\cdots,\mathbf{W}_{i,K_{\pi_i}}]$
According to *Schur complement*[1],

$$\begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}=\begin{bmatrix}\mathbf{\Sigma}_i+\mathbf{W}_i^T\mathbf{\Sigma}_{\pi_i}\mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T\mathbf{W}_i\\ \mathbf{W}_i^T\mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i}\end{bmatrix}^{-1} \tag{90}$$

Since $\mathbf{\Sigma}_{\pi_i}$ and $\mathbf{\Sigma}_i$ are positive definite, then

$$\mathbf{u}^T\begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\mathbf{u}=\left[\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\mathbf{u}\right]^T\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\left[\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\mathbf{u}\right]>0 \tag{91}$$

hold for any $\mathbf{u}$, thus,

$$\begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}$$

is positive definite.
Therefore,

$$\left(\mathbf{x}_{\pi_i}-\mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1}\left(\mathbf{x}_{\pi_i}-\mu_{\pi i}\right)+\left(\mathbf{x}_i-(\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\mathbf{x}_j)\right)^T \mathbf{\Sigma}_i^{-1}\left(\mathbf{x}_i-(\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\mathbf{x}_j)\right)\} \tag{92}$$

$$=\left[\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix}-\begin{bmatrix}\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}\right]^T\begin{bmatrix}\mathbf{\Sigma}_i+\mathbf{W}_i^T\mathbf{\Sigma}_{\pi_i}\mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T\mathbf{W}_i\\ \mathbf{W}_i^T\mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i}\end{bmatrix}^{-1}\left[\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix}-\begin{bmatrix}\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}\right] \tag{93}$$

Denote

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix}$$

$$\boldsymbol{\mu}=\begin{bmatrix}\mathbf{b}_i+\sum_{j\in x_{\pi_i}}\mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}$$

[1]https://en.wikipedia.org/wiki/Schur_complement

$$\Sigma = \begin{bmatrix} \mathbf{\Sigma}_i + \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} \mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T \mathbf{W}_i \\ \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i} \end{bmatrix}$$

Put everything together, we have

$$p(\mathbf{x}_i, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\} \tag{94}$$

Therefore,

$$(\mathbf{x}_i, \mathbf{x}_{\pi_i}) \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j \\ \mathbf{x}_{\pi_i} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_i + \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} \mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T \mathbf{W}_i \\ \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i} \end{bmatrix}\right)$$

Given this property, it follows easily by induction that the joint distribution over the graph G is multivariate Gaussian.