Q1.

REFERENCE :

<u>Entropy</u> : $H[x] = -\sum_x p(x) \log_2 p(x)$

$= \mathbb{E}_{p(x)}[-\log p(x)]$

<u>Joint entropy</u> : $H[x,y] = H[x] + H[y|x]$

$= H[y] + H[x|y]$

$= \mathbb{E}_{p(x,y)}[-\log p(x,y)]$

<u>conditional entropy</u> : $H[x|y] = \sum_x \sum_y p(x,y) \log p(x|y)$

$= \mathbb{E}_{p(x,y)}[-\log p(x|y)]$

<u>Mutual Info.</u> :

$I[x,y] = KL(p(x,y) \,||\, p(x)p(y))$

$= H[x] - H[x|y]$

<u>Cond. Mutual info</u> :

$I(x,y|z) = \mathbb{E}_{p(z)}[KL(p(x,y|z) \,||\, p(x|z)p(y|z))]$

$= H[x|z] - H[x|y,z]$

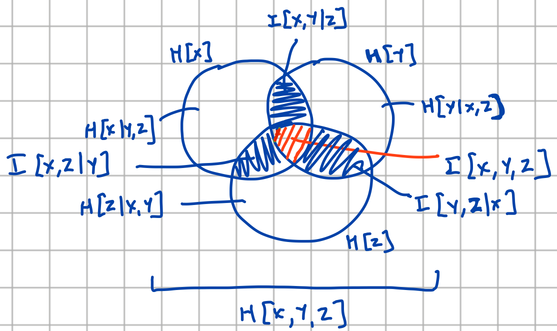
$= H[y|z] - H[y|x,z]$

a)

$$I(x,y,z) = I(x,y) - I(x,y|z)$$
$$= H[x] - H[x|y] - H[x|z] + H[x|y,z]$$
$$= I[x,z] - I[x,z|y] \quad \text{(collecting terms \& using definition)}$$
$$= I[x,z,y]$$

Hence proved.

b)



c)

$$x \to z \to y$$

Here, y is d-separated from x given x.
Thus, $x \perp\!\!\!\perp y | z$
and $I(x,y|z) = 0$

Now, we know, $I(x,y,z) = I(x,z,y)$

$$I[x,y] - \underbrace{I[x,y|z]}_{=0} = I[x,z] - I[x,z|y]$$

$$I[x,y] = I[x,z] - \underbrace{I[x,z|y]}_{\geq 0}$$

or, $I[x,y] \leq I[x,z]$

proved

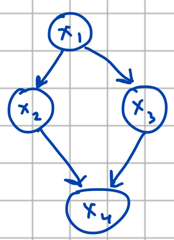
Q2.

a) Using BN factorization property : $p(x_1, \cdots, x_m) = \prod_{i=1}^{m} p(x_i \mid pa(x_i))$

$$p(x_1, x_2, x_3, x_4) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_1)\, p(x_4 \mid x_2, x_3)$$

// Ans.

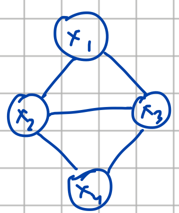
b)



1) $x_1 \perp\!\!\!\perp x_2 \mid \phi$ — False — direct edge between them.
2), 3), 4) →

5) $x_1 \perp\!\!\!\perp x_4 \mid \phi$ — False — non-colliders $\{x_2, x_3\}$ in path but not in condition. ∴ Unblocked paths
6) $x_1 \perp\!\!\!\perp x_4 \mid x_3$ — False — " " but x_2 not in cond. ∴ Unblocked paths.
7) $x_1 \perp\!\!\!\perp x_4 \mid x_2$ — False — " " " x_3 " " "
8) $x_1 \perp\!\!\!\perp x_4 \mid \{x_2, x_3\}$ — True — " " " in condition. All paths blocked
9) $x_2 \perp\!\!\!\perp x_3 \mid \phi$ — False — unblocked path from x_1.
10) $x_2 \perp\!\!\!\perp x_3 \mid x_1$ — True — all paths blocked.
11) $x_2 \perp\!\!\!\perp x_3 \mid x_4$ — False — all paths unblocked.
12) $x_2 \perp\!\!\!\perp x_3 \mid \{x_1, x_4\}$ — False — path from x_4 unblocked.

c)



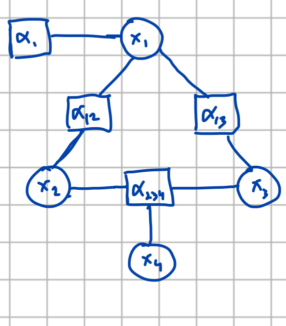
MRF

maximal cliques	corresponding potential func. as conditionals :
$\{x_1, x_2, x_3\}$	$\psi_{1,2,3}(x_1, x_2, x_3) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_1)$
$\{x_2, x_3, x_4\}$	$\psi_{2,3,4}(x_2, x_3, x_4) = p(x_4 \mid x_2, x_3)$

d)

1) $x_1 \perp\!\!\!\perp x_2 \mid \phi$ — False — all paths unblocked.
2) $x_1 \perp\!\!\!\perp x_2 \mid x_3$ — False — $x_2 - x_4 - x_3 - x_1$ & $x_2 - x_3 - x_1$ blocked BUT $x_1 - x_2$ unblocked.
3) $x_1 \perp\!\!\!\perp x_2 \mid x_4$ — False — 2 paths unblocked
4) $x_1 \perp\!\!\!\perp x_2 \mid \{x_3, x_4\}$ — False — $x_1 - x_2$ unblocked.
5) $x_1 \perp\!\!\!\perp x_4 \mid \phi$ — False — all paths unblocked
6) $x_1 \perp\!\!\!\perp x_4 \mid x_3$ — False — $x_1 - x_2 - x_4$ unblocked
7) $x_1 \perp\!\!\!\perp x_4 \mid x_2$ — False — $x_1 - x_3 - x_4$ unblocked
8) $x_1 \perp\!\!\!\perp x_4 \mid \{x_2, x_3\}$ — True — all blocked
9), 10), 11), 12) — False — $x_2 - x_3$ always unblocked

e)

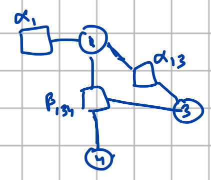


where,
$\alpha_1(x_1) = p(x_1)$
$\alpha_{12}(x_1, x_2) = p(x_2 \mid x_1)$
$\alpha_{13}(x_1, x_3) = p(x_3 \mid x_1)$
$\alpha_{234}(x_2, x_3, x_4) = p(x_4 \mid x_2, x_3)$

// Ans.

f) **Using Factor graphs**

$$p(x_1) = \sum_{x_4}\sum_{x_3}\sum_{x_2} \alpha_1(x_1)\,\alpha_{12}(x_1,x_2)\,\alpha_{13}(x_1,x_3)\,\alpha_{234}(x_2,x_3,x_4)$$

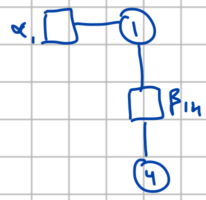


Step 1 : elimin. x_2

$$p(x_1) = \sum_{x_4}\sum_{x_3} \alpha_1(x_1)\,\alpha_{13}(x_1,x_3)\,\beta_{134}(x_1,x_3,x_4)$$

where, $\beta_{134} = \sum_{x_2} \alpha_{12}\,\alpha_{234}$

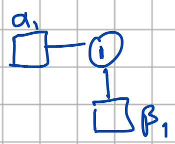
Step 2 : elim. x_3



$$p(x_1) = \sum_{4} \alpha_1(x_1)\,\beta_{1,4}(x_1,x_4)$$

where, $\beta_{14}(x_1,x_4) = \sum_{x_3} \alpha_{13}\,\beta_{134}$

Step 3 : elim. x_4



$$p(x_1) = \alpha_1(x_1)\,\beta_1(x_1)$$

where, $\beta_1(x_1) = \sum_{x_4} \beta_{14}$

Step	eliminate	variables induced	factors used	new factor
1	x_2	1, 2, 3, 4	α_{12} , α_{234}	β_{134}
2	x_3	1, 3, 4	α_{13}, β_{134}	β_{14}
3.	x_4	1, 4	β_{134}	β_1

Using MRF

$$p(x_1) = \sum_{x_2}\sum_{x_3}\sum_{x_4} p(x_1,x_2,x_3,x_4) = \underbrace{\sum_{x_4}\underbrace{\sum_{x_3}\underbrace{\sum_{x_2}\psi_{1,2,3}(x_1,x_2,x_3)\,\psi_{2,3,4}(x_2,x_3,x_4)}_{\tau_1(x_1,x_3,x_4)}}_{\tau_2(x_1,x_4)}}_{\tau_3(x_1)}$$

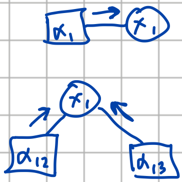
Step 1 : elim. x_2 $= \sum_{x_3}\sum_{x_4} \tau_1(x_1,x_3,x_4)$

Step 2 : elim x_3 $= \sum_{4} \tau_2(x_1,x_4)$

Step 3 : elim. x_4 $= \tau_3(x_1)$

Step	eliminate	variables induced	cliques used	new factor
1	x_2	1, 2, 3, 4	$\psi_{123}(x_1,x_2,x_3), \psi_{234}(x_2,x_3,x_4)$	$\tau_1(x_1,x_3,x_4)$
2	x_3	1, 3, 4	$\psi_{123}(x_1,x_2,x_3), \psi_{234}(x_2,x_3,x_4)$	$\tau_2(x_1,x_4)$
3.	x_4	1, 4	$\psi_{123}(x_1,x_2,x_3), \psi_{234}(x_2,x_3,x_4)$	$\tau_3(x_1)$

g)



$$\mu_{\alpha_1 \to x_1}(x_1) = f_1(x_1)$$

$$\mu_{\alpha_{12} \to x_1}(x_1) = f_{12}(x_1)\, \mu_{x_2 \to \alpha_{12}}(x_2)$$

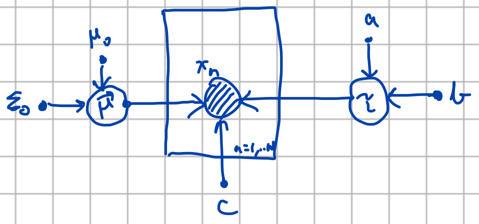
$$\mu_{13 \to x_1}(x_1) = f_{13}(x_1)\, \mu_{x_3 \to \alpha_{13}}(x_3)$$

$$\Bigg\} \quad p(x_1) = \frac{1}{Z}\, \mu_{1 \to x_1}(x_1)\, \mu_{12 \to x_1}(x_1)\, \mu_{13 \to x_1}(x_1)$$

// Ans.

$\boxed{\text{Q 3.}}$

a)



b) $$p(\mu, \tau, x) = p(x \mid \mu, \tau^{-1})\, p(\mu \mid \mu_0, \Sigma_0)\, p(\tau \mid a, b)$$

$$= \prod_{i=1}^{N} \mathcal{N}(x_n \mid \mu, \tau^{-1})\, \mathcal{N}(\mu \mid \mu_0, \Sigma_0)\, \text{gamma}(\tau \mid a, b)$$

// Ans.

c) Here, the posterior can be broken into conditionals as $p(\mu, \tau \mid x) \to p(\mu \mid \tau, x),\ p(\tau \mid \mu, x)$

d) f) Thus, in Gibbs sampling we first sample (say) $\underline{\mu^{new} \sim p(\mu \mid \tau, x)}$ & then, $\underline{\tau^{new} \sim p(\tau \mid \mu^{new}, x)}$ which is equivalent to sampling from their joint.

Sampling μ^{new} :

$$p(\mu \mid \tau, x) = \frac{p(\tau, x \mid \mu)\, p(\mu)}{p(\tau, x)} = \frac{p(x \mid \mu, \tau)\, \cancel{p(\tau \mid \mu)}\, p(\mu)}{p(x \mid \tau)\, \cancel{p(\tau)}}$$

[τ & μ are d-separated]

$$\propto p(x \mid \mu, \tau)\, p(\mu)$$

$$= \mathcal{N}(x \mid \mu, \tau^{-1} c)\, \mathcal{N}(\mu \mid \mu_0, \Sigma_0) \longrightarrow \exp\left(-\frac{1}{2}\left(\mu^T \underbrace{A}_{=\Sigma^{-1}} \mu - 2\mu^T \underbrace{A^{-1} B}_{=\mu} - B^T \mu\right]\right)$$

$$\text{or,} \quad \mu^{new} \sim \mathcal{N}\left(\mu \,\Big|\, \left(\tau N c + \frac{1}{\Sigma_0}\right)^{-1}, \left(\tau N c + \frac{1}{\Sigma_0}\right)\left(\tau c^{-1} \sum_n x_n + \Sigma^{-1} \mu_0\right)\right)$$ (completing the square)

// Ans.

Sampling τ^{new} :

Similarly $$p(\tau \mid x, \mu^{new}) \propto p(x \mid \mu^{new}, \tau)\, p(\tau)$$

$$= \mathcal{N}(x \mid \mu^{new}, \tau^{-1})\, \text{gamma}(\tau \mid a, b)$$

$$= \underline{\sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}(x - \mu_{new})^2\right)} \cdot \underline{\frac{b^a}{\Gamma(a)} \tau^{a-1} \exp(-b\tau)}$$

$$= \frac{b^a}{\Gamma(a)} \tau^{a-1+\frac{1}{2}} \exp\left(-\tau\left[\frac{(x - \mu^{new})^2}{2} + b\right]\right)$$

or, $\tau^{new} \sim \text{gamma}\left(\tau \mid a+\frac{1}{2}, \frac{(x-\mu^{new^2})}{2}+b\right)$

// Ans.

d) NO? if τ^{new} depends on μ^{new}?

g)

h) 1. $z_i \sim N(\mu, \sigma)$

2. $w_i = \dfrac{\tilde{p}(z_i)}{\tilde{q}(z_i)} = \dfrac{\text{gamma}(z_i \mid a, b)}{N(z_i \mid \mu, \sigma)} = \dfrac{z_i^{a-1} \exp(-b z_i)}{\exp\left(-\frac{1}{2\sigma}(z_i-\mu)^2\right)} = z_i^{a-1} \exp\left(-b z_i + \frac{1}{2\sigma}(z_i-\mu)^2\right)$

// Ans.

i) In general, to estimate

$$E[f(\tau)] = \int \text{gamma}(\tau \mid a, b)\, f(\tau)\, d\tau$$

we need to use the samples $x_i \sim N(x_i \mid \mu, \sigma)$ which are then used to calcul. w_i:

$$= \frac{\int p(x) \frac{f(x)}{q(x)} q(x)}{\int \frac{q(x)}{q(x)} p(x)\, dx} = \frac{\int q(x)\, w(x) f(\tau)\, d\tau}{\int q(x)\, w(x)\, dx} = \frac{E_{q(x)}[w(x) f(x)]}{E_{q(x)}[w(x)]}$$

$$\boxed{= \frac{\sum_i w(x_i) f(x_i)}{\sum_j w(x_j)}}$$

In our case, $p(x) = \text{gamma}(\tau \mid a, b)$
$q(x) = N(x_i \mid \mu, \sigma)$

Thus samples x_i from import. sampler are used to approx. the expectation.

j) Since, $w_i = \dfrac{\tilde{p}(x_i)}{\tilde{q}(x_i)}$ it is important that support of the distributions p & q are similar in all the regions to reduce variance. Thus, p & q should have similar forms & stats:

$E_p[x] = E_q[x] \Rightarrow \mu = \frac{a}{b}$ & $\sigma^2 = \frac{a}{b^2}$

p = gamma

q = Normal

k)

$\boxed{\text{Q4.}}$ (see Q4. of HW6 - exact same)