NATIONAL UNIVERSITY OF SINGAPORE

CS5340 - Uncertainty Modelling in AI

(Quiz 1, Semester 2 AY2022/23)

SOLUTIONS

Time Allowed: 1 hour

Instructions

- This is an open-book quiz. You may refer to any of the lecture slides and tutorials, or any of your own notes.
- You may *not* refer to any external online material or use any software to help you answer the questions. You may use a calculator.
- Please do not cheat; your answers *must* be your own. Do *not* collaborate with anyone else.
- Please put all your answers in Canvas.
- Read each question *carefully*. Don't get stuck on any one problem. The questions are *not* in any particular order of difficulty.
- Don't panic. The problems often look more difficult than they really are.
- Good luck!

Student Number.: ________________________________

Common Probability Distributions

Distribution (Parameters)	PDF/PMF
Normal (μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
Bernoulli (r)	$r^x(1-r)^{(1-x)}$
Categorical (π)	$\prod_{k=1}^{K} \pi_k^{x_k}$
Binomial (μ, N)	$\binom{N}{x}\mu^x(1-\mu)^{N-x}$
Poisson (λ)	$\frac{\lambda^x \exp[-\lambda]}{x!}$
Beta (α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
Gamma (a, b)	$\frac{1}{\Gamma(a)} b^a x^{a-1} \exp[-bx]$
Dirichlet $(\boldsymbol{\alpha})$	$\frac{\Gamma(\sum_k^K \alpha_k)}{\Gamma(\alpha_1)...\Gamma(\alpha_K)} \prod_{k=1}^{K} x_k^{\alpha_k-1}$
Multivariate Normal $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\frac{1}{(2\pi)^{D/2}\lvert\Sigma\rvert^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$
Uniform (a, b)	$\frac{1}{b-a}$

Note: $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ is the Gamma function.

1 True or False?

For the following questions, please answer TRUE or FALSE. **No justifications are needed. There is negative marking for this section (–0.5 points per wrong answer)**.

Grading Policy: 1 point for each correct answer.

Problem 1. [1 points] Given any three random variables X, Y and Z, then

$$\mathbb{V}[X + Y + Z] = \mathbb{V}[X] + \mathbb{V}[Y] + \mathbb{V}[Z]$$

Solution: False. X, Y, and Z may not be independent.

Problem 2. [1 points] Consider a Bayesian Network $B = (G, P)$ where G is a DAG and P is probability distribution that factorizes according to G. True or false: there always exists an *undirected* graph U such that U is an I-map for P.

Solution: True. Consider the definition of an I-map. A simple U that is an I-map for P is a fully connected graph.

Problem 3. [1 points] X and Y are independent random variables where $p(X) = \text{Normal}[\mu_X, \sigma_X^2]$ and $p(Y) = \text{Normal}[\mu_Y, \sigma_Y^2]$. True or false: The random variable $Z = \alpha X + Y$ has expectation $\mathbb{E}[Z] = \alpha\mu_X + \mu_Y$.

Solution: True

Problem 4. [1 points] X and Y are independent Bernoulli random variables. Specifically $X \sim \text{Bern}[\theta_X]$ and $Y \sim \text{Bern}[\theta_Y]$. True or false: The random variable $Z = X + Y$ is also Bernoulli distributed with expectation $\mathbb{E}[Z] = \theta_X + \theta_Y$.

Solution: False. Z is not Bernoulli since it has support [0,2].

Problem 5. [1 points] Scientists have discovered a new radioactive element that decays over time. The same scientists have estimated the half-life for the element to be 2.5 days. The uncertainty over the half-life estimate is an example of epistemic uncertainty.

Solution: True.

Problem 6. [1 points] If the empirical or sample covariance estimated via MLE from realizations of two Gaussian random variables X and Y is non-zero, then X and Y are *not* independent.

Solution: False.

Problem 7. [1 points] p is an exponential family distribution. True or False: we can always find a conjugate prior for p.

Solution: True.

Problem 8. [1 points] Assume an arbitrary distribution p over *continuous* variables X and Y with a corresponding Bayesian network G. Assume p factorizes according to G. It must be the case that G is an I-map for p.

Solution: True

Problem 9. [1 points] You are given a MRF with graph U and a positive distribution p that factorizes according to U. If two nodes X and Y are separated in U given observed variables Z, we can *always* conclude that $(X \perp Y|Z) \in I(p)$ where $I(p)$ is the independence set of p.

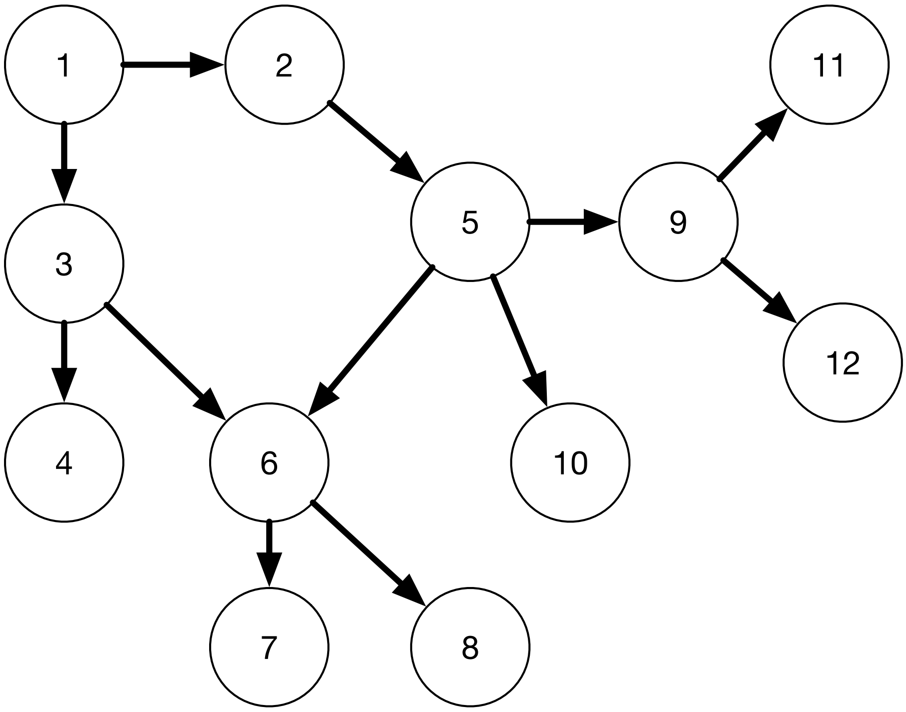
Solution: True.

Problem 10. [1 points] You are tasked to model the effectiveness of a particular treatment B, i.e., whether a patient who is given B will end up being cured. The effectiveness of B depends on the patient's age, denoted A. True or false: the marginal of $p(B) = \int_A p(B|A)p(A)dA$ can be modeled with a Bernoulli distribution.

Solution: True.

2 Blocking Nodes

You are given the following Bayesian Network.



2.1 d-separation Test

Each node represents a binary random variable. Check if each of the following 5 conditional independence assertions is TRUE or FALSE. **There is negative marking for the following 5 problems** (– 0.5 points per wrong answer).

Problem 11. [1 points] $(4 \perp 8|5)$

Solution: False

Problem 12. [1 points] $(4 \perp 10|\emptyset)$

Solution: False

Problem 13. [1 points] $(8 \perp 9|12)$

Solution: False

Problem 14. [1 points] $(6 \perp 7|4)$

Solution: False

Problem 15. [1 points] $(3 \perp 8 | \{2, 4\})$

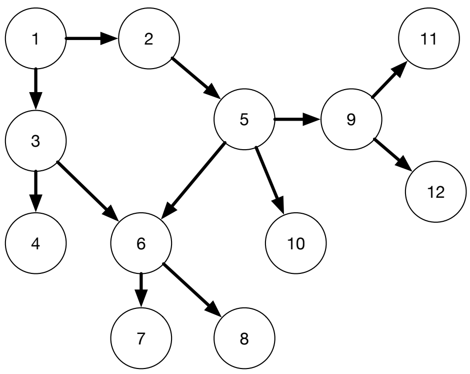
Solution: False

2.2 Choosing Nodes

For each of the following, **select *all* options** for $\mathcal{Z}$ that will render the conditional independence assertion true. **You must select all correct options and no wrong ones to earn full points for each question.** We have shown the graph on the right side of each question for ease of reference.

Problem 16. [2 points] $(4 \perp \{2, 10\} | \mathcal{Z})$

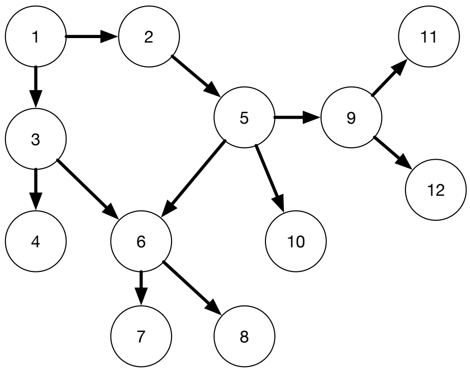
A. $\mathcal{Z} = \{1, 8, 9\}$
B. $\mathcal{Z} = \{3, 5\}$
C. $\mathcal{Z} = \{3\}$
D. $\mathcal{Z} = \{6, 7\}$
E. $\mathcal{Z} = \{12\}$
F. $\mathcal{Z} = \{3, 5, 7\}$
G. $\mathcal{Z} = \{3, 5, 12\}$



Solution: C

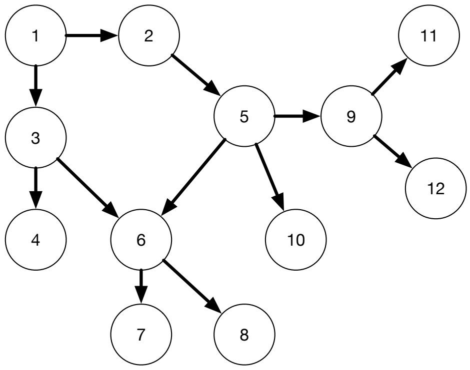
Problem 17. [2 points] $(1 \perp \{10, 11\} | \mathcal{Z})$

A. $\mathcal{Z} = \{4, 8\}$
B. $\mathcal{Z} = \{3, 9\}$
C. $\mathcal{Z} = \{9, 12\}$
D. $\mathcal{Z} = \{3, 9, 12\}$
E. $\mathcal{Z} = \{3, 12\}$
F. $\mathcal{Z} = \{5, 8\}$
G. $\mathcal{Z} = \{7, 12\}$



Solution: A, B, C, D, F

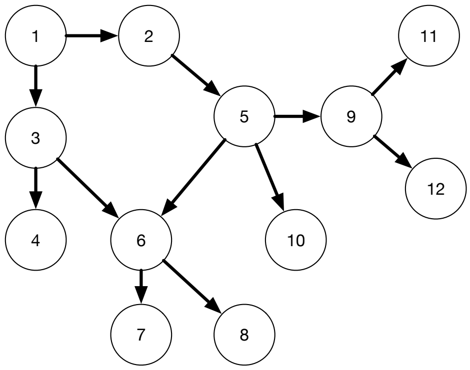
Problem 18. [2 points] $(4 \perp 12 | \mathcal{Z})$



A. $\mathcal{Z} = \{3\}$
B. $\mathcal{Z} = \{9\}$
C. $\mathcal{Z} = \{3, 9\}$
D. $\mathcal{Z} = \{5, 9\}$
E. $\mathcal{Z} = \{3, 5\}$
F. $\mathcal{Z} = \{3, 2\}$
G. $\mathcal{Z} = \{7, 9\}$

Solution: A, B, C, D, F, G

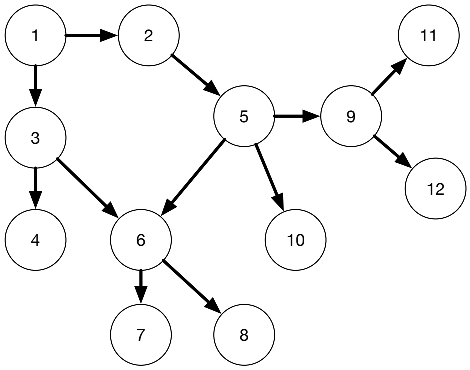
Problem 19. [2 points] $(2 \perp 11 | \mathcal{Z})$



A. $\mathcal{Z} = \{3\}$
B. $\mathcal{Z} = \{4\}$
C. $\mathcal{Z} = \{4, 5, 6\}$
D. $\mathcal{Z} = \{8, 9\}$
E. $\mathcal{Z} = \{5, 8\}$
F. $\mathcal{Z} = \{12\}$
G. $\mathcal{Z} = \{8, 9, 10\}$

Solution: D, E, G.

Problem 20. [2 points] $(1 \perp 2 | \mathcal{Z})$



A. $\mathcal{Z} = \{3\}$
B. $\mathcal{Z} = \{4\}$
C. $\mathcal{Z} = \{5\}$
D. $\mathcal{Z} = \{8\}$
E. $\mathcal{Z} = \{9\}$
F. $\mathcal{Z} = \{12\}$
G. $\mathcal{Z} = \{8, 9, 10\}$

Solution: D, E, F, G

3 Noisy Regression

You have some data $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} = \{(x_i, y_i)\}_{i=1}^N$ where $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$, i.e., both are real numbers. You have developed a linear regression model,

$$y = \mathbf{w}^\top \begin{bmatrix} x \\ 1 \end{bmatrix} + \epsilon \tag{1}$$

where $\mathbf{w} = [w_1, w_2]^\top$ and $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. In other words, assume a linear model with i.i.d. Gaussian noise. As usual, the noise is unobserved.

Problem 21. [1 points] Given the setup above, what is the conditional distribution $p(y|x, \mathbf{w})$?

A. $p(y|x, \mathbf{w}) = \mathcal{N}(y|w_1 x + w_2, \sigma_n^2)$
B. $p(y|x, \mathbf{w}) = \mathcal{N}(y|w_1 x, \sigma_n^2)$
C. $p(y|x, \mathbf{w}) = \mathcal{N}(y|w_2 x + w_1, \sigma_n^2)$
D. $p(y|x, \mathbf{w}) = \mathcal{N}(y|w_2 x + w_1, 1)$
E. $p(y|x, \mathbf{w}) = \mathrm{Cat}[w_1 x, w_2]$
F. $p(y|x, \mathbf{w}) = \mathrm{Bern}[w_1 x]$
G. $p(y|x, \mathbf{w}) = \mathrm{Bern}[w_1 x + w_2]$

Solution: A. $p(y|x, \mathbf{w}) = \mathcal{N}(y|w_1 x + w_2, \sigma_n^2)$

After training the model on the data, you observe that the model's performance is poor. You discover that the noise variable ϵ isn't distributed as you initially assumed above. In fact, with probability $\alpha = 0.9$, the noise is distributed $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$, but with probability $(1 - \alpha)$, $\epsilon \sim \mathcal{N}(0, \beta^2)$ where $\beta^2 > \sigma_n^2$. This process produces outliers in your data.

Let us introduce a new random variable z which captures which noise variance applies. Assume that the values for σ_n^2 and β^2 are known, but z is unobserved.

Problem 22. [2 points] Which of the following is an appropriate model for z and ϵ given our updated information?

A. $p(z) = \alpha^z (1-\alpha)^{(1-z)}$ and $p(\epsilon|z) = \begin{cases} \mathcal{N}(0, \sigma_n^2) & \text{if } z = 1 \\ \mathcal{N}(0, \beta^2) & \text{otherwise} \end{cases}$
B. $p(z) = \alpha^z (1-\alpha)^{(1-z)}$ and $p(\epsilon|z) = \begin{cases} \mathcal{N}(0, \beta^2) & \text{if } z = 1 \\ \mathcal{N}(0, \sigma_n^2) & \text{otherwise} \end{cases}$
C. $p(z) = \alpha^z (1-\alpha)^{(1-z)}$ and $p(\epsilon|\alpha) = \begin{cases} \mathcal{N}(0, \sigma_n^2) & \text{if } z = 1 \\ \mathcal{N}(0, \beta^2) & \text{otherwise} \end{cases}$
D. $p(z) = \alpha^z (1-\alpha)^{(1-z)}$ and $p(\epsilon|\alpha) = \begin{cases} \mathcal{N}(0, \beta^2) & \text{if } z = 1 \\ \mathcal{N}(0, \sigma_n^2) & \text{otherwise} \end{cases}$
E. $p(z) = \mathcal{N}(\alpha, 1)$ and $p(\epsilon|\alpha) = \begin{cases} \mathcal{N}(0, \sigma_n^2) & \text{if } z \geq \alpha \\ \mathcal{N}(0, \beta^2) & \text{otherwise} \end{cases}$
F. $p(z) = \mathcal{N}(\alpha, 1)$ and $p(\epsilon|\alpha) = \begin{cases} \mathcal{N}(0, \beta^2) & \text{if } z \geq \alpha \\ \mathcal{N}(0, \sigma_n^2) & \text{otherwise} \end{cases}$
G. $p(z) = \mathcal{N}(\alpha, 1)$ and $p(\epsilon|z) = \begin{cases} \mathcal{N}(0, \sigma_n^2) & \text{if } z \geq \alpha \\ \mathcal{N}(0, \beta^2) & \text{otherwise} \end{cases}$

H. $p(z) = \mathcal{N}(\alpha, 1)$ and $p(\epsilon|z) = \begin{cases} \mathcal{N}(0, \beta^2) & \text{if } z \geq \alpha \\ \mathcal{N}(0, \sigma_n^2) & \text{otherwise} \end{cases}$

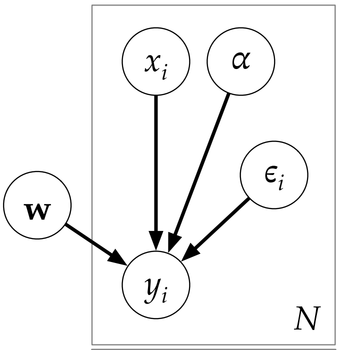
Solution: A. $p(z) = \alpha^z(1-\alpha)^{(1-z)}$ and $p(\epsilon|z) = \begin{cases} \mathcal{N}(0, \sigma_n^2) & \text{if } z = 1 \\ \mathcal{N}(0, \beta^2) & \text{otherwise} \end{cases}$

Problem 23. [2 points] As before, we will assume a linear model but the noise distribution has been updated based on our new understanding, i.e.,

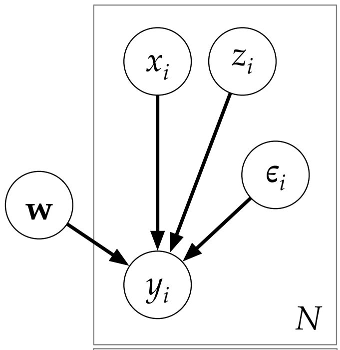
$$y = \mathbf{w}^\top \begin{bmatrix} x \\ 1 \end{bmatrix} + \epsilon$$

where now ϵ depends on z. Which of the DAGs best represents our updated model on the dataset $\mathcal{D}$? Note: to reduce clutter, we have not shown the known parameters σ_n^2 and β^2 and not shaded observed nodes. (Options G and H shown on the next page)

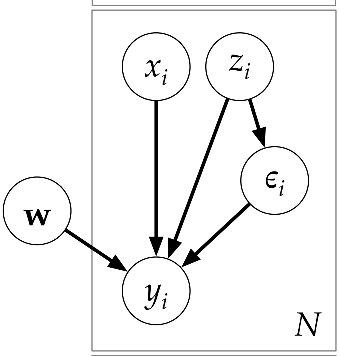
A.



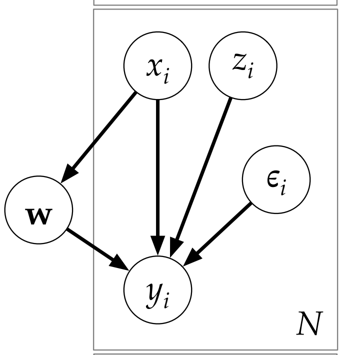
D.



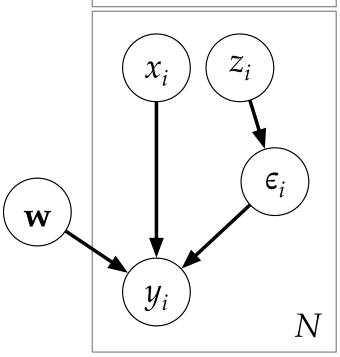
B.



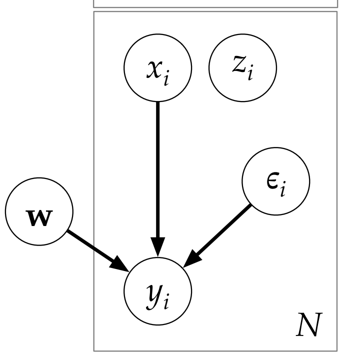
E.



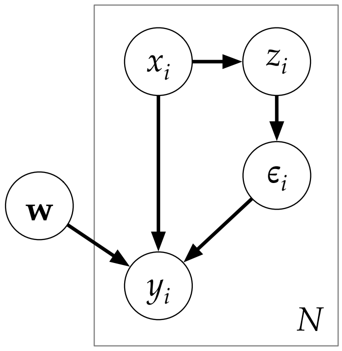
C.



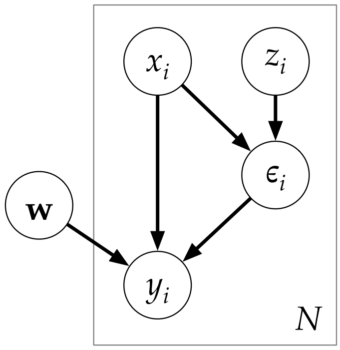
F.



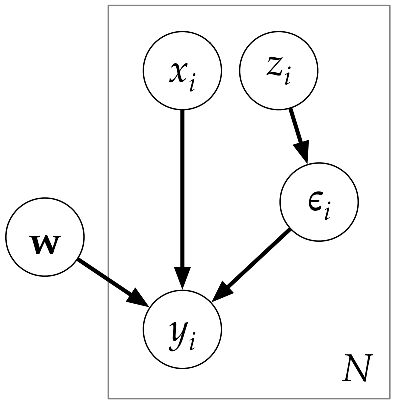
G.



H.



Solution: C.



Problem 24. [4 points] If we wish to obtain the maximum likelihood estimate for $\mathbf{w}$, which of the following functions $L(\mathbf{w};\alpha)$ should we *minimize*?

A. $L(\mathbf{w};\alpha) = -\prod_i \left[\alpha \log \mathcal{N}(y_i|w_1x_i + w_2, \sigma_n^2) + (1-\alpha)\log \mathcal{N}(y_i|w_1x_i + w_2, \beta^2)\right]$
B. $L(\mathbf{w};\alpha) = -\sum_i \left[\alpha \log \mathcal{N}(y_i|w_1x_i + w_2, \sigma_n^2) + (1-\alpha)\log \mathcal{N}(y_i|w_1x_i + w_2, \beta^2)\right]$
C. $L(\mathbf{w};\alpha) = -\prod_i \log[\alpha^{\mathcal{N}(y_i|w_1x_i+w_2,\sigma_n^2)} + (1-\alpha)^{\mathcal{N}(y_i|w_1x_i+w_2,\beta^2)}]$
D. $L(\mathbf{w};\alpha) = -\sum_i \log[\alpha^{\mathcal{N}(y_i|w_1x_i+w_2,\sigma_n^2)} + (1-\alpha)^{\mathcal{N}(y_i|w_1x_i+w_2,\beta^2)}]$
E. $L(\mathbf{w};\alpha) = -\prod_i \log[\alpha\mathcal{N}(y_i|w_1x_i + w_2, \sigma_n^2) + (1-\alpha)\mathcal{N}(y_i|w_1x_i + w_2, \beta^2)]$
F. $L(\mathbf{w};\alpha) = -\sum_i \log[\alpha\mathcal{N}(y_i|w_1x_i + w_2, \sigma_n^2) + (1-\alpha)\mathcal{N}(y_i|w_1x_i + w_2, \beta^2)]$

Solution: F. $L(\mathbf{w};\alpha) = -\sum_i \log[\alpha\mathcal{N}(y_i|w_1x_i + w_2, \sigma_n^2) + (1-\alpha)\mathcal{N}(y_i|w_1x_i + w_2, \beta^2)]$

Problem 25. [2 points] Suppose now that we do not know the value for α, but we have some external information that it should be around 0.8. We can place a Beta prior over α and perform maximum a posteriori (MAP) estimation: which of the following functions L_{MAP} should we minimize? In all the following options, assume that $L(\mathbf{w};\alpha)$ corresponds to the correct answer to the previous question. **Pick the best answer among the options below.**

A. $L_{\text{MAP}}(\mathbf{w},\alpha) = L(\mathbf{w};\alpha)$
B. $L_{\text{MAP}}(\mathbf{w},\alpha) = L(\mathbf{w};\alpha) - 7\log\alpha - \log(1-\alpha)$

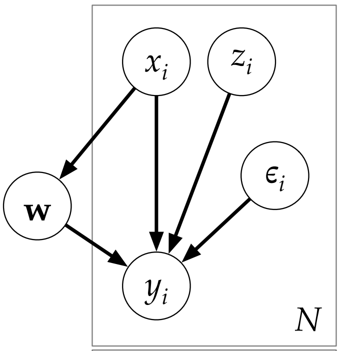
C. $L_{\text{MAP}}(\mathbf{w}, \alpha) = L(\mathbf{w}; \alpha) - \log \alpha - 7 \log(1 - \alpha)$
D. $L_{\text{MAP}}(\mathbf{w}, \alpha) = L(\mathbf{w}; \alpha) - 7 \log \alpha - \log(1 - \alpha)$ subject to $0 \leq \alpha \leq 1$
E. $L_{\text{MAP}}(\mathbf{w}, \alpha) = L(\mathbf{w}; \alpha) - \log \alpha - 7 \log(1 - \alpha)$ subject to $0 \leq \alpha \leq 1$
F. $L_{\text{MAP}}(\mathbf{w}, \alpha) = L(\mathbf{w}; \alpha) - \lambda[7 \log \alpha + \log(1 - \alpha)]$ where λ is a Lagrangian multiplier
G. $L_{\text{MAP}}(\mathbf{w}, \alpha) = L(\mathbf{w}; \alpha) - \lambda[\log \alpha + 7 \log(1 - \alpha)]$ where λ is a Lagrangian multiplier

Solution: D.

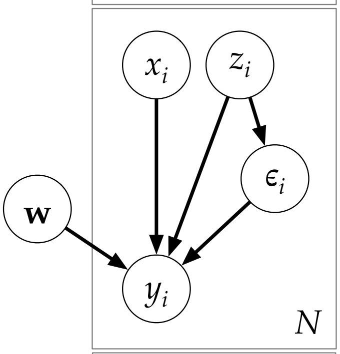
3.1 Input Dependence

Problem 26. [2 points] Now, consider that you learn that the underlying noise distribution also depends on x. In particular, the *probability* of outliers (samples with noise variance β^2) is higher for larger x. Which of the following graphs best represents our new model where the noise level (σ_n^2 or β^2) depends on the input x?

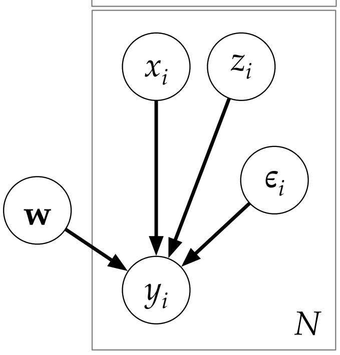
A.



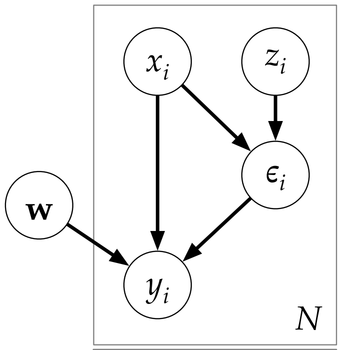
B.



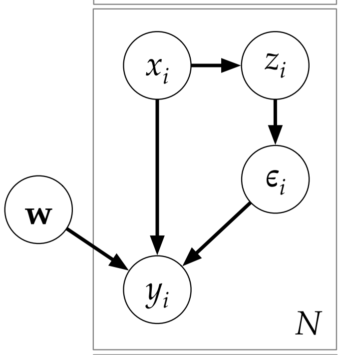
C.



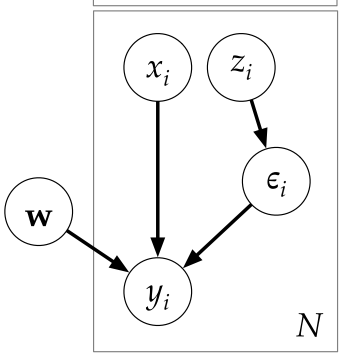
D.

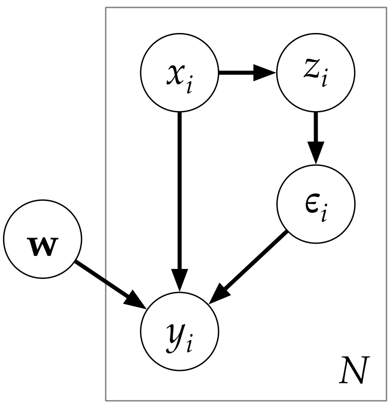


E.



F.





Solution: E.

Problem 27. [2 points] Which of the following distributions for $p(z|x)$ should we apply given our model above? Choose the best option among the following. Assume $\zeta > 0$ to be a parameter of the distribution.

A. $p(z|x,\zeta) = \frac{1}{1+\exp(\zeta x)}$
B. $p(z|x,\zeta) = \frac{1}{1+\exp(-\zeta x)}$
C. $p(z|x,\zeta) = \zeta x$
D. $p(z|x,\zeta) = -\zeta x$
E. $p(z|x,\zeta) = x^{\zeta}$
F. $p(z|x,\zeta) = x^{-\zeta}$
G. $p(z|x,\zeta) = \text{Bern}[\zeta]$
H. $p(z|x,\zeta) = \text{Bern}[\zeta x]$
I. $p(z|x,\zeta) = \mathcal{N}(z|x,\zeta)$

Solution: B. $p(z|x,\zeta) = \frac{1}{1+\exp(-\zeta x)}$. This is equivalent to $p(z|x,\zeta) = \text{Bern}\left[\frac{1}{1+\exp(-\zeta x)}\right]$

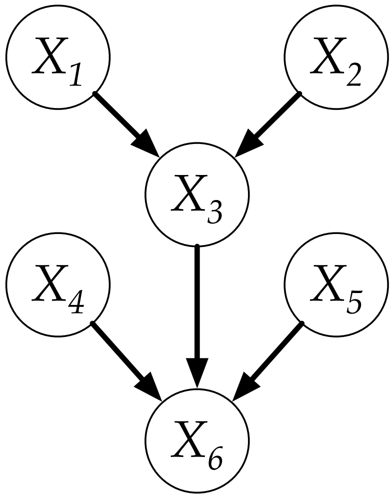
Problem 28. [1 points] We wish to obtain a MAP estimate for the ζ parameter above. Which of the following priors can we use as $p(\zeta)$?

A. Bernoulli
B. Multinomial
C. Gamma
D. Normal
E. Uniform[-10,10]
F. Dirichlet

Solution: C. Gamma.

4 Expectations

You are given the following Bayesian Network.



Note the following facts about this Bayesian network:

- $p(X_1) = \mathcal{N}(\mu_1 = 2, \sigma_1^2 = 1)$
- $p(X_2) = \mathcal{N}(\mu_2 = 3, \sigma_2^2 = 2)$
- $X_3 = X_1 + 2X_2$
- $p(X_4) = \text{Bern}[0.2]$
- $p(X_5) = \text{Binomial}[10, 0.5]$
- $X_6 = X_3 + X_4 + 2X_5$

Recall: the mean of a Bernoulli random variable with parameter p is just p and its variance is $p(1-p)$. The mean of a Binomial random variable with parameters (n, p) is given by np and its variance is given by $np(1-p)$.

Problem 29. [1 points] What is $p(X_3)$?

A. $\mathcal{N}(\mu_3 = 8, \sigma_3^2 = 3)$
B. $\mathcal{N}(\mu_3 = 8, \sigma_3^2 = 5)$
C. $\mathcal{N}(\mu_3 = 8, \sigma_3^2 = 9)$
D. $\mathcal{N}(\mu_3 = 5, \sigma_3^2 = 3)$
E. $\mathcal{N}(\mu_3 = 5, \sigma_3^2 = 5)$
F. $\mathcal{N}(\mu_3 = 5, \sigma_3^2 = 9)$
G. $\mathcal{N}(\mu_3 = 11, \sigma_3^2 = 5)$
H. $\mathcal{N}(\mu_3 = 11, \sigma_3^2 = 3)$
I. $\mathcal{N}(\mu_3 = 11, \sigma_3^2 = 9)$
J. None of the above

Solution: C. $\mathcal{N}(\mu_3 = 8, \sigma_3^2 = 9)$

Problem 30. [2 points] What is the expectation of X_6, i.e., $\mathbb{E}[X_6]$?

A. 13.2
B. 15.2
C. 16.2
D. 18.2
E. 21.2
F. 22.2
G. None of the above

Solution: D. 18.2

Problem 31. [2 points] What is the variance of X_6, i.e., $\mathbb{V}[X_6]$?

A. 6.16
B. 8.16
C. 10.16
D. 13.16
E. 15.16
F. 19.16
G. None of the above

Solution: F. 19.16

Problem 32. [2 points] Conditioned upon $X_3 = 2$, what is the expectation of X_5, i.e., $\mathbb{E}[X_5]$?

A. 0.5
B. 2
C. 2.5
D. 4.5
E. 5
F. 5.5
G. None of the above

Solution: E. 5. Conditioning on X_3 has no impact on the expectation of X_5 since X_6 is unobserved.

Problem 33. [2 points] Conditioned upon $X_3 = 2$, what is the variance of X_5, i.e., $\mathbb{V}[X_5]$?

A. 0.5
B. 2
C. 2.5
D. 4.5
E. 5
F. 5.5
G. None of the above

Solution: C. 2.5. Observing X_3 has no impact on the variance of X_5 since X_6 is unobserved.

Problem 34. [2 points] Conditioned upon $X_3 = 3$ and $X_6 = 6$, what is the expectation of X_5, i.e., $\mathbb{E}[X_5]$?

A. 0.2
B. 1.0
C. 1.4
D. 2.6
E. 3.2
F. 3.4
G. None of the above

Solution: B. 1.0

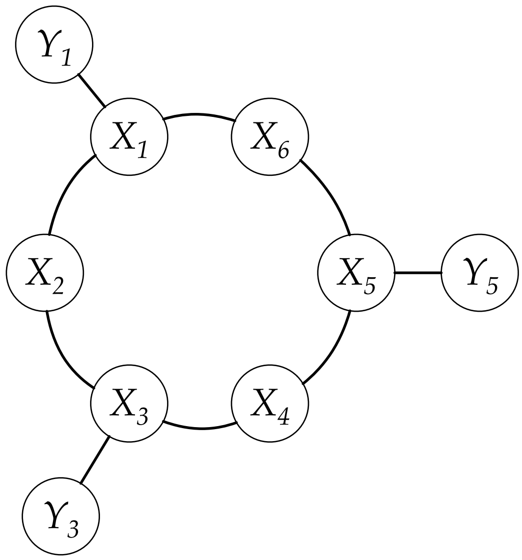
Problem 35. [2 points] Conditioned upon that $X_3 = 3$, and $X_6 = 6$, what is the variance of X_5, i.e., $\mathbb{V}[X_5]$?

A. 0.20
B. 0.25
C. 0.71
D. 1.00
E. 1.42
F. 2.84
G. None of the above

Solution: None of the above.

5 Ring of Probability

You are given the following Markov Random Field, which forms a ring with three spokes. All random variables are binary random variables.



5.1 Markov Blanket

A Markov Blanket is the set of nodes such that if given random variable X_i is conditioned upon its Markov blanket $MB(X_i)$, X_i is independent of all other nodes in the graph. Here, we are interested in finding the **Markov Boundary** which is the *minimal* Markov Blanket, i.e., the Markov Boundary is the smallest Markov blanket. In other words, if you remove any node from the Markov Boundary, the set is no longer a Markov blanket.

For each of the following questions in this subsection, **select *all* nodes** that are in the Markov Boundary . **You must select all correct options and no wrong ones to earn full points for each question.**

Problem 36. [1 points] Which of the following nodes are in the Markov Boundary of X_1?

A. X_1
B. X_2
C. X_3
D. X_4
E. X_5
F. X_6
G. Y_1
H. Y_3
I. Y_5

Solution: Y_1, X_2, X_6.

Problem 37. [1 points] Which of the following nodes are in the Markov Boundary of X_4?

A. X_1
B. X_2
C. X_3
D. X_4
E. X_5
F. X_6
G. Y_1
H. Y_3
I. Y_5

Solution: X_3, X_5.

Problem 38. [1 points] Which of the following nodes are in the Markov Boundary of Y_5?

A. X_1
B. X_2
C. X_3
D. X_4
E. X_5
F. X_6
G. Y_1
H. Y_3
I. Y_5

Solution: X_5.

Problem 39. [1 points] Which of the following nodes are in the Markov Boundary of X_2?

A. X_1
B. X_2
C. X_3
D. X_4
E. X_5
F. X_6
G. Y_1
H. Y_3
I. Y_5

Solution: X_1, X_3.

5.2 Energy Design

In this subsection, suppose that the joint probability distribution is given by:

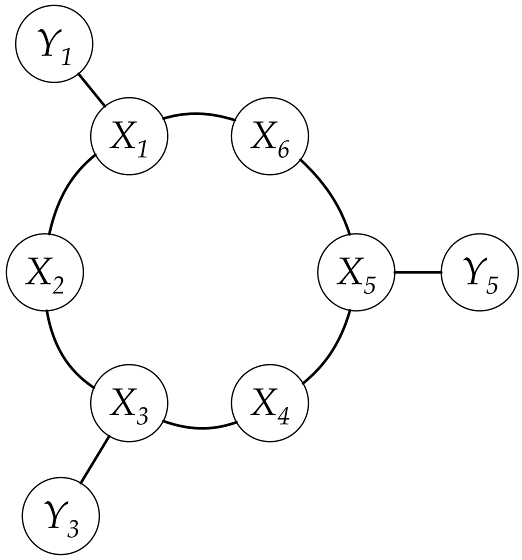
$$p(\mathcal{X}, \mathcal{Y}) = \frac{1}{Z} \exp[-E(\mathcal{X}, \mathcal{Y})] \tag{2}$$

where we have used the shorthand $\mathcal{X}$ to refer to the set of all X_i in the graph (and likewise for $\mathcal{Y}$). All random variables either take on a value of -1 or 1.

We adopt the energy function,

$$E(\mathcal{X}, \mathcal{Y}) = \sum_i f(X_i, Y_i) + \sum_{(k,l) \in U} g(X_k, X_l) \tag{3}$$

where U is the set of all undirected edges in the MRF.

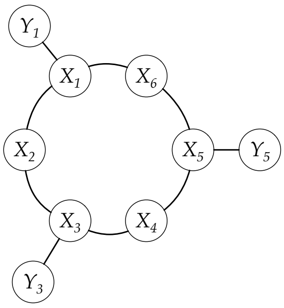


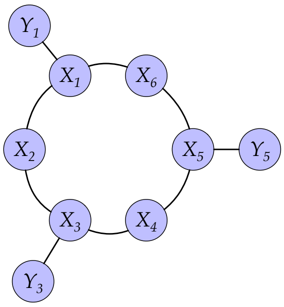
Problem 40. [2 points] Suppose we parameterize the energy function with the following.

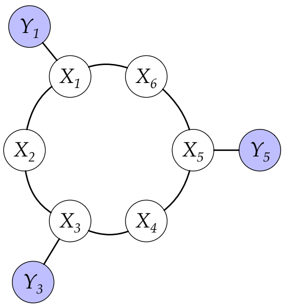
- $f(X_i, Y_i) = -\alpha X_i Y_i$ where $\alpha > 0$
- $g(X_k, X_l) = \beta X_k X_l$ where $\beta > 0$

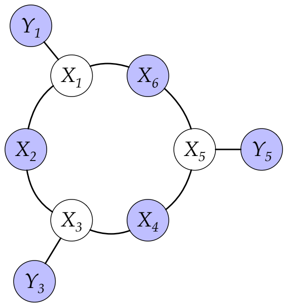
and we condition upon $Y_1 = -1, Y_3 = -1$.

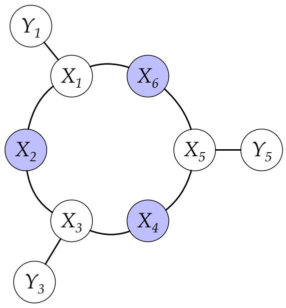
The graphs below show various realizations of the random variables. Shaded nodes have value 1 and unshaded nodes have value -1. Among the graphs shown below, select the one which has the highest probability.

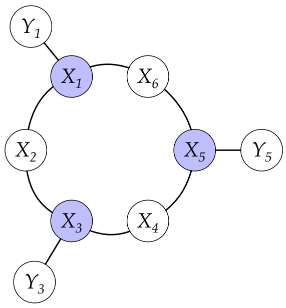
A.


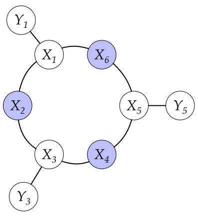
B.


C.


D.


E.


F.


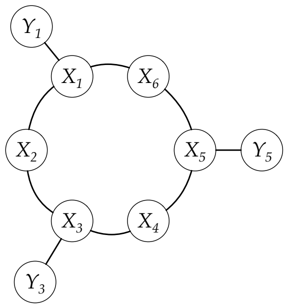
Solution: E.


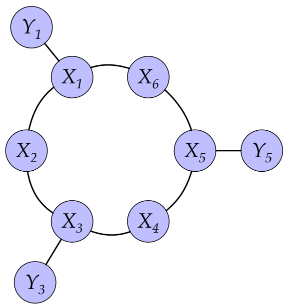
Problem 41. [2 points] Suppose we parameterize the energy function with the following.

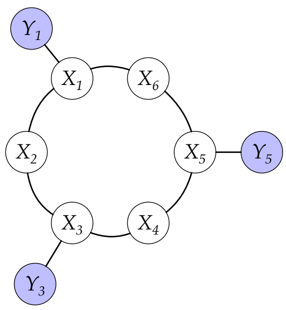
- $f(X_i, Y_i) = \alpha X_i Y_i$ where $\alpha > 0$
- $g(X_k, X_l) = \beta X_k X_l$ where $\beta > 0$

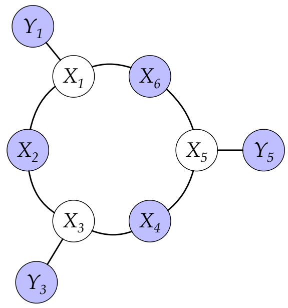
and we condition upon $Y_1 = -1, Y_5 = -1$.

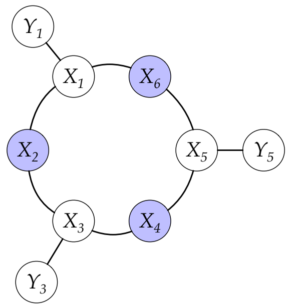
The graphs below show various realizations of the random variables. Shaded nodes have value 1 and unshaded nodes have value -1. Among the graphs shown below, select the one which has the highest probability.

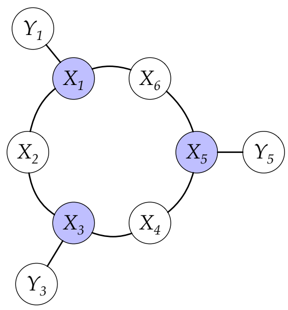
A.


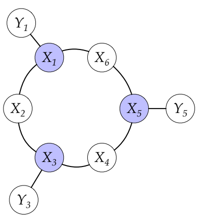
B.


C.


D.


E.


F.


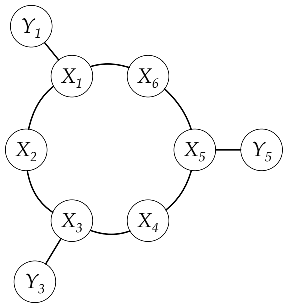
Solution: F.


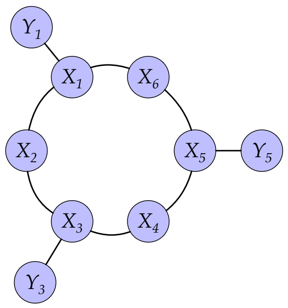
Problem 42. [2 points] Suppose we parameterize the energy function with the following.

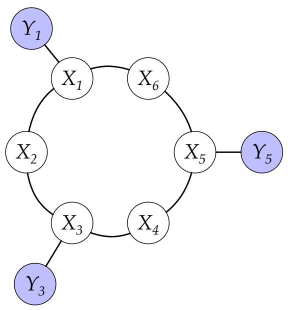
- $f(X_i, Y_i) = \alpha X_i Y_i$ where $\alpha > 0$
- $g(X_k, X_l) = -\beta X_k X_l$ where $\beta > 0$

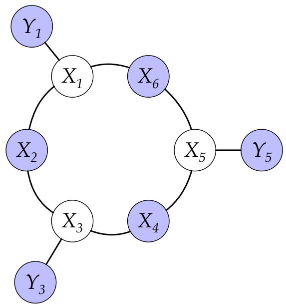
and we condition upon $Y_1 = 1, Y_5 = 1$.

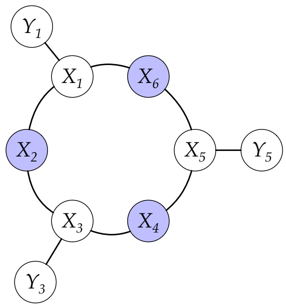
The graphs below show various realizations of the random variables. Shaded nodes have value 1 and unshaded nodes have value -1. Among the graphs shown below, select the one which has the highest probability.

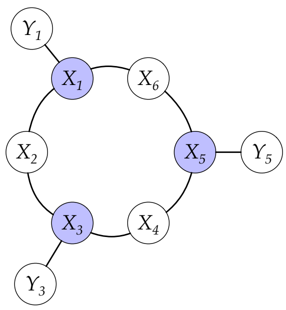
A.


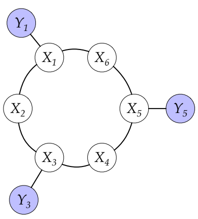
B.


C.


D.


E.


F.


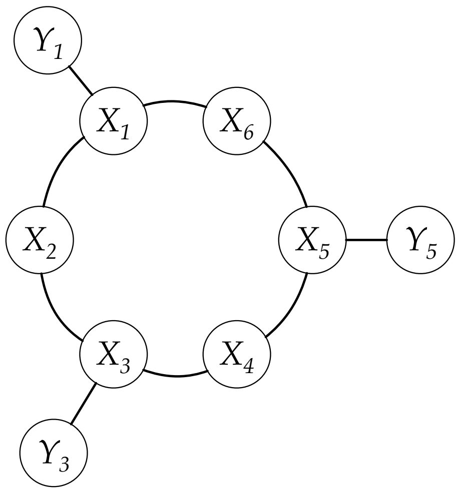
Solution: C.


5.3 Conditional Computations

In this subsection, suppose the joint probability distribution is given by:

$$p(\mathcal{X}, \mathcal{Y}) = \prod_i f(X_i, Y_i) \prod_{(k,l)\in U} g(X_k, X_l) \tag{4}$$

where we have used the shorthand $\mathcal{X}$ to refer to the set of all X_i in the graph (and likewise for $\mathcal{Y}$) and U is the set of all undirected edges in the MRF. The functions f and g are given by the tables below.



X_i	Y_i	$f(X_i, Y_i)$
-1	-1	5
-1	1	1
1	-1	1
1	1	3

X_k	X_l	$g(X_k, X_l)$
-1	-1	1
-1	1	4
1	-1	4
1	1	1

Problem 43. [2 points] What is $p(Y_1 = 1|X_1 = -1)$?

A. 1/6
B. 5/6
C. 1/10
D. 5/10
E. 1/5
F. 4/5
G. 1/10
H. 4/10
I. None of the above.

Solution: A. 1/6

Problem 44. [2 points] What is $p(Y_5 = -1|X_5 = -1)$?

A. 1/6
B. 5/6
C. 1/10
D. 5/10
E. 1/5
F. 4/5
G. 1/10
H. 4/10
I. None of the above.

Solution: B. 5/6

Problem 45. [2 points] What is $p(X_1 = -1|X_2 = -1, X_6 = 1)$?

A. 1/5
B. 2/5
C. 3/5
D. 4/5
E. 1/6
F. 2/6
G. 3/6
H. 4/6
I. 5/6
J. None of the above.

Solution: C. 3/5

Problem 46. [2 points] What is $p(X_5 = 1|X_4 = 1, X_6 = -1)$?

A. 1/5
B. 2/5
C. 3/5
D. 4/5
E. 1/6
F. 2/6
G. 3/6
H. 4/6
I. 5/6
J. None of the above.

Solution: B. 2/5