EECS 126 Probability and Random Processes — UC Berkeley, Fall 2022
Kannan Ramchandran — Dec 16, 2022

Final

Last Name	First Name	SID

Left Neighbor Full Name	**Right Neighbor** Full Name	**Room Number**

***Rules*.**

- **Please bubble in your answers FULLY and write all numerical answers clearly. Answers that are not legible or clearly bubbled in may not get credit.**
- You have 70 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $70 \cdot X\%$ time on the exam).
- This exam is not open book. You may reference three double-sided handwritten sheets of paper. No calculator or phones allowed.
- Collaboration with others is strictly prohibited. If you are caught cheating, you will receive a 0 on the final and will face disciplinary consequences.
- Write in your SID on every page to receive 1 point.

Problem	points earned	out of
SID		1
Problem 1		14
Problem 2		14
Problem 3		10
Problem 4		10
Problem 5		14
Problem 6		19
Problem 7		12
Problem 8		14
Problem 9		18
Total		126

1 Random Cut of Random Graph [14 points]

Recall that a *cut* of a graph G is a subset of vertices $T \subseteq G$, and an edge (i, j) is said to be *across* the cut T if and only if exactly only one of its endpoints i or j belongs to T.

Let $G \sim \mathcal{G}(100, 1/4)$ be an Erdős–Rényi random graph on 100 vertices, in which each edge appears independently with probability 1/4. We construct a *random cut* of G by selecting each vertex of G with probability 1/3. Find the expected number of edges that cross this random cut of the random graph G.

- ○ 400.
- ○ 450.
- ○ 500.
- ○ 550.
- ○ 600.
- ○ None of the above.

The answer is $\boxed{550}$.

We write $n = 100$, $p = \frac{1}{4}$, and $q = \frac{1}{3}$. Any particular cut of size k has $k(n-k)$ possible crossing edges; each edge appears w.p. p, so the expected number of crossing edges is $k(n-k)p$. Then, by the law of total expectation, we condition on the size of the random cut to find

$$\begin{aligned}
&\sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} \cdot k(n-k)p \\
&= n(n-1)pq(1-q) \sum_{k=0}^{n} \frac{(n-2)!}{(k-1)!(n-k-1)!} q^{k-1}(1-q)^{n-k-1} \\
&= n(n-1)pq(1-q) \\
&= 100 \times 99 \times \frac{1}{4} \times \frac{2}{9}.
\end{aligned}$$

Alternatively, let $C_{i,j}$ be the event that (i, j) is a crossing edge. Note that $\mathbb{P}(C_{i,j}) = \mathbb{P}(C_{1,2}) = \frac{1}{9}$: an edge between i and j exists w.p. $\frac{1}{4}$, and the two vertices are on opposite sides of the cut w.p. $2(\frac{1}{3})(\frac{2}{3}) = \frac{4}{9}$, independently of the edge existing. Then, by the linearity of expectation,

$$\mathbb{E}\left(\sum_{i<j} \mathbf{1}_{C_{i,j}}\right) = \sum_{i<j} \mathbb{E}(\mathbf{1}_{C_{i,j}}) = \binom{100}{2} \mathbb{P}(C_{1,2}) = \frac{100 \times 99}{2} \times \frac{1}{9} = 550.$$

2 Vogel im Käfig [14 points]

A bird lives on the integers $\mathbb{Z}$. It starts at 0 at time 0. At each time step, it jumps one step left or right with probability $\frac{1}{2}$ each. In other words, if X_n is its position at time n, then $X_{n+1} = X_n + 1$ w.p. $\frac{1}{2}$ and $X_n - 1$ w.p. $\frac{1}{2}$. If p_n is the probability that the bird is outside of the interval $[-\sqrt{n}, \sqrt{n}]$ at time n, find $\lim_{n\to\infty} p_n$.

Give a **numerical** answer to two decimal places. You may use these following values of $\Phi(\cdot)$, the standard normal CDF: $\Phi(-2) \approx 0.02$, $\Phi(-1) \approx 0.16$, $\Phi(-0.5) \approx 0.31$, $\Phi(0.5) \approx 0.69$, $\Phi(1) \approx 0.84$, $\Phi(2) \approx 0.98$.

The answer is $\boxed{0.32}$.

Let us write $X_{n+1} = X_n + Y_n$, where Y_n are i.i.d. Rademacher random variables, which are ± 1 with probability $\frac{1}{2}$ each. Observing that $X_n = \sum_{k=0}^{n-1} Y_k$, by the Central Limit Theorem,

$$\frac{1}{\sqrt{n}} X_n = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} Y_k \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

Note that Rademachers have mean 0 and variance 1. Then

$$p_n = \mathbb{P}(|X_n| > \sqrt{n}) = \mathbb{P}\left(\left|\frac{X_n}{\sqrt{n}}\right| > 1\right) = \mathbb{P}\left(\left|\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} Y_k\right| > 1\right) \to \mathbb{P}(|Z| > 1) = 2\Phi(-1)$$

for $Z \sim \mathcal{N}(0,1)$.

3 Poisson Arrivals [10 points]

Consider a Poisson process $(N_t)_{t\geq 0}$ with rate $\lambda = 1$. For $i \in Z^+$, let T_i be the time of the ith arrival.

(a) Find $\mathbb{E}[T_3 \mid N(1) = 2]$

By the memoryless property, $\mathbb{E}[T_3 \mid N(1) = 2] = 1 + \mathbb{E}[T_1] = 1 + \lambda^{-1} = 2$

(b) Find $\mathbb{E}[T_2 \mid T_3 = 1]$. Format your answer in reduced fraction form.

By part (b), T_2 is the maximum of two uniform random variables between 0 and s. Thus, if $0 \leq x \leq s$,

$$F_{T_2|T_3=s}(x) = \Pr(T_2 \leq x \mid T_3 = s) = \left(\frac{x}{s}\right)^2$$

and

$$f_{T_2|T_3=s}(x) = \frac{2x}{s^2}\mathbf{1}\{0 \leq x \leq s\}.$$

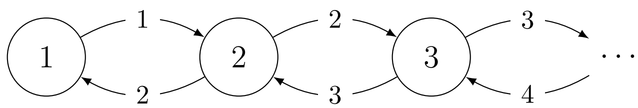
Finally,

$$\mathbb{E}[T_2 \mid T_3 = s] = \int_0^s \frac{2x^2}{s^2}\,\mathrm{d}x = \frac{2s}{3}.$$

Therefore, $A = 2$, $B = 3$

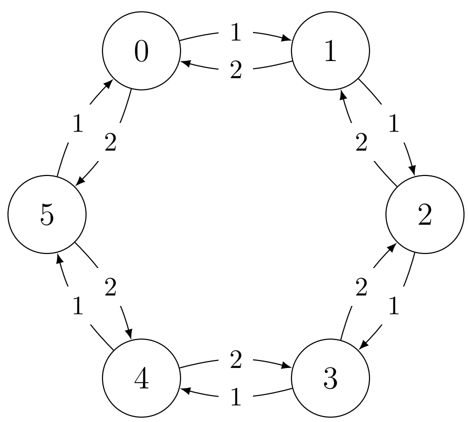
4 Reversible CTMCs [10 points]

For each of the following transition *rate* diagrams, select true if it describes a *reversible* continuous-time Markov chain and false otherwise.



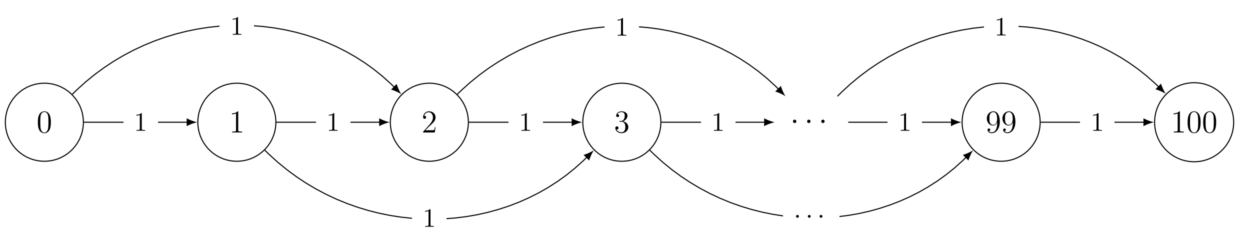
◯ True ◯ False

False. The harmonic series diverges, so no probability distribution can satisfy detailed balance. Alternatively, the embedded jump chain is the classic symmetric reflected random walk on $\mathbb{Z}$, which does not have a stationary distribution by its null recurrence.



◯ True ◯ False

False. The unique stationary distribution, which is uniform, fails to satisfy detailed balance.



◯ True ◯ False

True. The stationary distribution with all mass in state 100 satisfies detailed balance vacuously.

5 German Tank Problem [14 points]

A bin contains a set of N balls with serial numbers $1, 2, \ldots, N$, where N is unknown. The goal is to estimate N based on a sample of the serial numbers. (Recall that this is the same setting as the German Tank problem discussed in class.) Suppose X is a randomly sampled serial number, i.e. X can take any of the serial numbers from 1 to N with equal probability; if we have a sample of $n \leq N$ serial numbers $X_1, X_2, \ldots, X_n$ sampled at random and without replacement from the bin, and our observed numbers for $n = 4$ are the sequence {33, 7, 100, 44}.

(a) What is the Maximum Likelihood Estimate (MLE) of N given our observed sequence {33, 7, 100, 44} for $n = 4$?

We compute the likelihood as

$$\mathbb{P}(X_1 = 33, X_2 = 7, X_3 = 100, X_4 = 44 \mid N = n) = \mathbf{1}\{\max\{x_1, x_2\} \leq n\} \frac{1}{\binom{n}{4}}.$$

We seek to find

$$\arg\max_n \mathbf{1}\{\max\{x_1, x_2, x_3, x_4\} \leq n\} \frac{1}{\binom{n}{4}} = \arg\min_{n \geq \max\{x_1, x_2, x_3, x_4\}} \binom{n}{4}$$

This function is increasing over that interval, so the MLE estimate is at the left endpoint of the interval, $\boxed{100}$.

(b) It is possible to show that the expected value of the MLE of N given $n \leq N$ random samples without replacement is given by $\frac{n(N+1)}{n+1}$. Use this to construct an unbiased estimator of N for $n = 4$ given the observed sequence $\{33, 7, 100, 44\}$. (Recall that an unbiased estimator, $\hat{X}$ of a random variable X is such that $\mathbb{E}[\hat{X}]$ is equal to $\mathbb{E}[X]$.)

From the problem statement, we know that

$$\mathbb{E}[\hat{N}_{\text{MLE}}] = \frac{n(N+1)}{n+1}$$

$$\frac{n+1}{n}\mathbb{E}[\hat{N}_{\text{MLE}}] = N + 1$$

$$\frac{n+1}{n}\mathbb{E}[\hat{N}_{\text{MLE}}] - 1 = N$$

Which shows that the LHS of the equation is the unbiased estimator $\mathbb{E}[\hat{N}_{\text{unbiased}}]$ of N.

$$\hat{N}_{\text{unbiased}} = \frac{n+1}{n}\hat{N}_{\text{MLE}} - 1.$$

Plugging in $n = 4$, we have:

$$\hat{N}_{\text{unbiased}} = \frac{5}{4}\hat{N}_{\text{MLE}} - 1 = \frac{5}{4}(100) - 1 = \boxed{124}.$$

(c) In a Bayesian setting, suppose the prior distribution on N is Geometric(p) with $p = 0.01$. What is the MAP estimate for N given $n = 4$ and the observed sequence $\{33, 7, 100, 44\}$?

We use the posterior

$$\begin{aligned}
&\arg\max_n \mathbb{P}(N = n \mid X_1 = 33, X_2 = 7, X_3 = 100, X_4 = 44) \\
=&\arg\max_n \mathbb{P}(X_1 = 33, X_2 = 7, X_3 = 100, X_4 = 44 \mid N = n)\,\mathbb{P}(N = n) \\
=&\arg\max_n \mathbf{1}\{n \geq 100\}\frac{1}{\binom{n}{4}}(1-p)^{n-1}p \\
=&\arg\max_n (n-1)\log(1-p) - \log(n) - \log(n-1) - \log(n-2) - \log(n-3).
\end{aligned}$$

We notice that this is a decreasing function of n. Therefore, the value n that maximizes the objective function is the smallest n satisfying $n \geq \max\{33, 7, 100, 44\}$, which is $\boxed{100}$.

6 Hypothesis Testing [19 points]

Recall the optimization problem solved by the Neyman-Pearson rule:

$$\max_{\hat{X}} \text{PCD} := \Pr\left(\hat{X} = 1 \mid X = 1\right)$$
$$\text{subj. to PFA} := \Pr\left(\hat{X} = 1 \mid X = 0\right) \leq \beta$$

for some fixed $\beta \in [0, 1]$.

(a) Suppose that $Y \mid \{X = 0\}$ and $Y \mid \{X = 1\}$ have the same distribution (e.g. Y is independent of X). Which best describes the relationship between the PFA and PCD for the Neyman-Pearson rule?

- ◯ PFA $\geq$ PCD, but we cannot determine if equality holds without knowing the distribution of Y and/or β.
- ◯ PFA $=$ PCD.
- ◯ PFA $\leq$ PCD, but we cannot determine if equality holds without knowing the distribution of Y and/or β.
- ◯ PFA $= 1 -$ PCD.
- ◯ We cannot determine without further information.

(b) is correct. As Y has the same distribution under $\{X = 0\}$ and $\{X = 1\}$, we conclude that the likelihood ratio $L(y)$ will take on the constant value 1. This will cause the optimal decision rule to be outputting Bernoulli(γ) for $\gamma = \beta$, which will cause the PFA and PCD to both be exactly γ.

(b) Suppose that $Y \mid \{X = 0\} \sim N(0, 1)$ and $Y \mid \{X = 1\} \sim N(0, 2)$. We solve for the Neyman-Pearson rule with the constraint that our PFA cannot exceed $\beta = 0.3$. Which of the following best describes the shape of the likelihood ratio $L(y)$?

- ◯ monotonically increasing
- ◯ monotonically decreasing
- ◯ increasing and then decreasing
- ◯ decreasing and then increasing
- ◯ none of the above

(d) is correct. The simplest solution is to graph the densities of $Y \mid \{X = 1\}$ versus $Y \mid \{X = 0\}$. The former distribution has a shorter peak and heavier tails, and both distributions are symmetric about 0. we conclude that the likelihood ratio should have a minimum $y = 0$ and increase as $|y|$ increases, which is consistent with the likelihood rate decreasing towards 0 and then increasing again.

Otherwise, we can solve for $L(y)$ analytically by dividing the normal pdfs:

$$\begin{aligned} L(y) &= \frac{(1/\sqrt{2\pi \cdot 2})\exp(-y^2/4)}{(1/\sqrt{2\pi})\exp(-y^2/2)} \\ &= \frac{1}{\sqrt{2}}\exp\left(\frac{y^2}{4}\right) \end{aligned}$$

Using the fact that $z \mapsto e^z$ is monotonic, we conclude that $L(y)$ is decreasing and then increasing.

Now suppose that we have the following conditional distributions for Y:

$$Y \mid \{X = 0\} = \begin{cases} 0 & \text{w.p. } 1/6 \\ 1 & \text{w.p. } 1/3 \\ 2 & \text{w.p. } 1/2 \end{cases}$$

$$Y \mid \{X = 1\} \sim \text{Uniform}\{0, 1, 2\}$$

Compute the Neyman-Pearson decision rule $\hat{X}(Y)$ given the constraint that the PFA cannot exceed $\beta = 2/5$. Then, compute the following values. Format your answers as fractions in reduced form.

(c) $\Pr\left(\hat{X} = 1 \mid Y = 0\right)$

(d) $\Pr\left(\hat{X} = 1 \mid Y = 1\right)$

(e) $\Pr\left(\hat{X} = 1 \mid Y = 2\right)$

We find

$$\Pr\left(\hat{X} = 1 \mid Y = 0\right) = 1$$
$$\Pr\left(\hat{X} = 1 \mid Y = 1\right) = 7/10$$
$$\Pr\left(\hat{X} = 1 \mid Y = 2\right) = 0$$

Observe that the likelihood function is a decreasing function of y, so we can greedily add values into our acceptance region. Always accepting for $Y = 0$ gives a PFA of $1/6$, which is below our constraint of $\beta = 2/5$, while always accepting for $Y \in \{0, 1\}$ gives a PFA of $1/2$, which is above our constraint of $\beta = 2/5$. Thus, we choose to always accept when $Y = 0$, and then accept with probability $\gamma = (2/5 - 1/6)/(1/3) = 7/10$ when $Y = 1$, and always reject when $Y = 2$.

7 ABC's [12 points]

Let X, Y, and Z be jointly Gaussian random variables with covariance matrix

$$\begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 2 \\ 0 & 2 & 3 \end{bmatrix}$$

and mean vector $[0, 126, 0]$. We can write $\mathbb{E}[Y|X, Z]$ as $a + bX + cZ$.

Compute a.

$$\mathbb{E}[\mathbb{E}[Y \mid X, Z]] = \mathbb{E}[a + bX + cZ] = a = \mathbb{E}[Y].$$

Therefore, $a = 126$.

Compute b.

By the orthogonality principle, we know that $Y - \mathbb{E}[Y \mid X, Z]$ is orthogonal to X. This yields

$$\begin{aligned} \mathbb{E}[(Y - 126 - bX - cZ)X] &= \mathbb{E}[(Y - 126)X] - b\,\mathbb{E}[X^2] - c\,\mathbb{E}[ZX] \\ &= 2 - 3b \\ &= 0, \end{aligned}$$

where we use the fact that X and Z are uncorrelated. Thus, $b = \frac{2}{3}$.

Compute c.

Similarly to the previous part,

$$\begin{aligned} \mathbb{E}[(Y - 126 - bX - cZ)Z] &= \mathbb{E}[(Y - 126)Z] - b\,\mathbb{E}[XZ] - c\,\mathbb{E}[Z^2] \\ &= 2 - 3c \\ &= 0. \end{aligned}$$

Therefore, $c = \frac{2}{3}$. Alternatively, we can use symmetry to see that this answer should be the same as the previous part.

8 Hilbert's 25th Problem [14 points]

We will work in the Hilbert space of real-valued random variables $\mathcal{H}$, equipped with the usual inner product $\langle X, Y\rangle = \mathbb{E}(XY)$. Determine whether the following statements are true or false in general.

(a) If X is orthogonal to 1, then X is zero-mean.

◯ True ◯ False

True. $X \perp 1$ means that $\langle X, 1\rangle = \mathbb{E}(X \cdot 1) = \mathbb{E}(X) = 0$.

(b) The norm $\|X\| = \sqrt{\langle X, X\rangle}$ always equals the standard deviation $\sigma_X = \sqrt{\text{var}(X)}$.

◯ True ◯ False

False. $\sqrt{\mathbb{E}(X^2)} = \sigma_X$ iff $\mathbb{E}(X^2) = \text{var}(X)$, or $\mathbb{E}(X) = 0$, which is not true in general.

(c) X and Y are independent if and only if they are orthogonal.

◯ True ◯ False

False. None of the Hilbert space conditions are strong enough to imply independence.

(d) Suppose $\text{cov}(X, Y) \neq 0$. Then $\mathsf{proj}_{\{X,Y\}}(Z) \neq \mathsf{proj}_X(Z) + \mathsf{proj}_Y(Z)$.

◯ True ◯ False

False. Correlated random variables might still be orthogonal, in which case equality holds. For example, let X, Y be jointly Gaussian with $\mu_X = 1$, $\mu_Y = -1$, and covariance 1.

9 Estimate the Right Option [18 points]

(a) Which of the following does the expectation of a random variable always minimize (if the expectation exists)?

- ◯ mean squared error, i.e. $\arg\min_{x\in\mathbb{R}} \mathbb{E}\left[(X-x)^2\right] = \mathbb{E}[X]$.
- ◯ mean absolute error, i.e. $\arg\min_{x\in\mathbb{R}} \mathbb{E}\left[|X-x|\right] = \mathbb{E}[X]$.
- ◯ probability of error, i.e. $\arg\min_{x\in\mathbb{R}} \mathbb{P}(X \neq x) = \mathbb{E}[X]$.
- ◯ none of the above

(b) Which of the following is true when we estimate X from Y?

- ◯ The MMSE is always strictly better than the LLSE in terms of mean squared error.
- ◯ If X and Y are both Gaussian, the MMSE equals the LLSE.
- ◯ We can still use the MMSE and the LLSE if the relationship between X and Y is unknown.
- ◯ None of the above.

(c) Which of the following is the estimation error of $\mathbb{L}[X|Y]$ always orthogonal to?

- ◯ all functions of Y
- ◯ all linear functions of Y but not all functions of Y in general
- ◯ all linear functions of X
- ◯ none of the above

(d) Suppose that Y and Z are zero-mean random variables. Decide which of the following statements are true in general. Select **all** correct options.

- ☐ $\mathbb{L}(X \mid Y, Z) = \mathbb{L}(X \mid Y) + \mathbb{L}(X \mid Z)$.
- ☐ If Y and Z are orthogonal, then $\mathbb{L}(X \mid Y, Z) = \mathbb{L}(X \mid Y) + \mathbb{L}(X \mid Z)$.
- ☐ If $X = aY + bZ$, then $\mathbb{L}(X \mid Y, Z) = \mathbb{L}(X \mid Y) + \mathbb{L}(X \mid Z)$.
- ☐ None of the above.

(e) You want to determine the value of $X \sim \mathcal{N}(0,1)$. However, your measurements are imprecise: you observe Y_1 and Y_2, where each Y_i is X plus some independent noise $Z_i \sim \mathcal{N}(0,1)$. Find the MMSE estimate of X given $Y_1 = -4$ and $Y_2 = 10$.

- ◯ 0
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 6

(1) $\boxed{\text{a is correct.}}$ This directly comes from the Hilbert projection theorem as $\arg\min_{x\in\mathbb{R}} \mathbb{E}\left[(X-x)^2\right]$ is the projection of X onto 1, which is $\langle X, 1\rangle = \mathbb{E}[X]$.

b is incorrect. Consider the RV X that takes 0 w.p. 0.2, 1 w.p. 0.2, and 10 w.p. 0.6. $\mathbb{E}[X] = 6.2$ has a mean absolute error of 4.56. However, 10 has a mean absolute error of 3.8, which is lower.

c is incorrect. Consider the same example as above, where the mean 6.2 has a probability of error of 1, but 10 has a probability of error 0.4, which is lower.

d is incorrect since a is correct.

(2) a is incorrect. If X and Y are JG, the estimation error of MMSE and LLSE are equal, so it's not strictly better.

b is incorrect as they need to be JG.

c is incorrect because we need to know $\text{cov}(X, Y)$ in the LLSE calculation and need to know the expectation of $X|Y$ to get the MMSE.

$\boxed{\text{d is correct.}}$ since all other choices are indeed incorrect.

(3) $\boxed{\text{b is correct.}}$ This directly comes from lecture. (Note that even if we interpret the term linear to be distinguished from affine, b is still the only correct statement, since orthogonality to all $aY + b$ implies orthogonality to all aY.)

(4) None of the above are correct. The usual orthogonal update requires X, Y, Z all zero-mean and Y, Z orthogonal. Here, we explore the importance of these conditions in making sure $\mathbb{L}(X \mid Y) + \mathbb{L}(X \mid Z)$ *avoids redundant projections onto* 1 *or onto* Y.

a. Either counterexample below shows that option a is false as well.
b. Take $X = Y + Z + 1$. Then $\mathbb{L}(X \mid Y, Z) = X$, while $\mathbb{L}(X \mid Y) + \mathbb{L}(X \mid Z) = X + 1$ double counts the mean of X.
c. X is now zero-mean, but suppose $X = Y = Z$ is not constant, so Y is not orthogonal to Z. Then $\mathbb{L}(X \mid Y, Z) = X$, but $\mathbb{L}(X \mid Y) + \mathbb{L}(X \mid Z) = 2X$.

(5) The answer is $\boxed{2}$. Observe that X, Y_1, Y_2 are jointly Gaussian, so $\mathbb{E}(X \mid Y_1, Y_2) = \mathbb{L}(X \mid Y_1, Y_2)$. Let us orthogonalize the given information:

$$\tilde{Y}_1 = Y_1 - \mathbb{E}(Y_1) = Y_1$$
$$\tilde{Y}_2 = Y_2 - \mathbb{L}(Y_2 \mid \tilde{Y}_1) = Y_2 - \frac{\text{cov}(\tilde{Y}_1, Y_2)}{\text{var}(\tilde{Y}_1)}\tilde{Y}_1 = Y_2 - \frac{1}{2}Y_1.$$

Note that $\text{var}(\tilde{Y}_2) = \text{var}(\frac{1}{2}X + Z_2 - \frac{1}{2}Z_1) = \frac{1}{4} + 1 + \frac{1}{4}$ by centeredness and independence.

Now, by orthogonal updates,

$$\begin{aligned}
\mathbb{L}(X \mid Y_1, Y_2) = \mathbb{L}(X \mid \tilde{Y}_1, \tilde{Y}_2) &= \mathbb{L}(X \mid \tilde{Y}_1) + \mathbb{L}(X \mid \tilde{Y}_2) \\
&= \frac{\operatorname{cov}(X, \tilde{Y}_1)}{\operatorname{var}(\tilde{Y}_1)}\tilde{Y}_1 + \frac{\operatorname{cov}(X, \tilde{Y}_2)}{\operatorname{var}(\tilde{Y}_2)}\tilde{Y}_2 \\
&= \frac{1}{2}Y_1 + \frac{1}{3}\left(Y_2 - \frac{1}{2}Y_1\right) \\
&= \frac{Y_1 + Y_2}{3}.
\end{aligned}$$

Alternatively, we see that $\mathbb{L}(X \mid Y_1, Y_2)$ is a constant multiple of $Y_1 + Y_2$ by symmetry. The constant factor $a = \frac{1}{3}$ is determined by the orthogonality principle:

$$\begin{aligned}
\mathbb{E}((X - \hat{X}) \cdot Y_i) &= \mathbb{E}([(1 - 2a)X - a(Z_1 + Z_2)] \cdot Y_i) \\
&= (1 - 2a)\,\mathbb{E}(X^2) - a\,\mathbb{E}(Z_i^2) \\
&= 1 - 3a \\
&= 0.
\end{aligned}$$