Final

Last Name	First Name	SID

- You have 10 minutes to read the exam and 150 minutes to complete this exam.
- The maximum you can score is 130, but 100 points is considered perfect.
- The exam is not open book, but you are allowed to consult the cheat sheet that we provide. No calculators or phones. No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- Show all work to get any partial credit.
- Take into account the points that may be earned for each problem when splitting your time between the problems.

Problem	points earned	out of
Problem 1		25
Problem 2		20
Problem 3		20
Problem 4		15
Problem 5		25
Problem 6		10
Problem 7		15
Total		100 (+30)

Problem 1: Answer these questions briefly but clearly. [25]

(a) Random Dog Walk [5]

A dog walks on the integers, possibly reversing direction at each step with probability $p = 0.1$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:

$$(x_0, x_1, x_2, x_3, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, \ldots).$$

True / False: The sequence of random variables $(X_n)_{n \geq 0}$ is a Markov chain.

(b) MCMC [5]

Let $\{0, 1\}$ be the state space of a Markov chain. Suppose that we try to simulate from the distribution $\pi = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ via Metropolis-Hastings using the proposal distribution:

$$p(x, y) = \begin{cases} \frac{1}{3}, & y = x \\ \frac{2}{3}, & y = 1 - x \end{cases}, \qquad x = 0, 1.$$

In other words, at state x we propose state y with probability $p(x, y)$. What should we set the acceptance probabilities $A(x, y)$ to be in order to make the algorithm correct (i.e., the stationary distribution of the chain is π)?

(c) Book Shop [5]

A small book shop has room for at most two customers. Potential customers arrive at a Poisson rate of ten customers per hour; they enter if there is room and are turned away, never to return, otherwise. The bookie serves the admitted customers in order, requiring an exponentially distributed time of mean four minutes per customer.

Suppose there are two customers in the store. Find the expected amount of time until the store is empty again.

(d) Random Graphs [5]

Let G be a random graph generated using the $\mathcal{G}(n, p)$ model. For what value of p is the expected number of cliques of five vertices in G equal to 1? (A *clique* on n vertices is a subset of n vertices such that for every pair of vertices in the clique, the edge between the vertices exists in the graph.)

(e) Graphical MMSE [5]

Let (X, Y) be a pair of random variables which are uniformly distributed over the region shown in Figure 1.

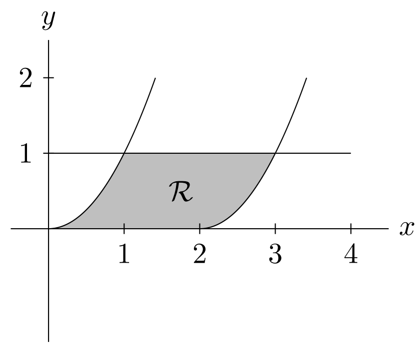

Figure 1: The region $\mathcal{R}$ is enclosed by the curves: $y = 0,\ y = 1,\ y = x^2,\ y = (x-2)^2$.

Find $\mathbb{E}[X \mid Y]$.

Problem 2: Waiting in Line for Infinity War (no spoilers) [20]

At 6:30 PM, there are 9 people waiting in line outside of the theater to watch *Avengers: Infinity War*. People arrive and wait in line according to a Poisson process with rate 2 per minute.

(a) [5] Starting from 6:30 PM, the theater admits exactly one person every minute (i.e., at 6:31 PM, 6:32 PM, 6:33 PM, etc.). What is the expected number of people remaining in line the instant after 6:40 PM?

(b) [5] Suppose we observe the time Y at which the third new person arrives in line. What is the MLE estimate for the time X at which the fourth new person arrives in line? (Assume that X and Y are measured in minutes since 6:30 PM.)

(c) [5] Starting from 6:40 PM, the theater stops letting in more people. The people in line become fed up: each person in line, independently, waits an Exponential(1) amount of time and then leaves the line. (People are still arriving in line according to the Poisson process of rate 2 per minute. If a new person arrives, he or she also waits an Exponential(1) amount of time before leaving the line.) After a long period of time, what is the average number of people in line?

(d) [5] Meanwhile, inside the theater, the movie has already begun. Thanos appears on the big screen according to a Poisson process of rate 1 per minute. On each of his appearances, independently with probability 1/3 he makes the audience scream. An observer outside the theater observes S, the number of times that the audience screams in the first 30 minutes. What is the LLSE estimate of T, the number of times that Thanos appeared on the big screen in the first 30 minutes, given S?

Problem 3: Hypothesis Testing [20]

(a) [10] We want to test two hypotheses; we have prior knowledge that $X \sim \text{Bernoulli}(2/3)$. In both cases, our observation Y has the Laplace distribution, but with different mean and shape. In particular

$$f_{Y|X}(y \mid 0) = \frac{1}{2}\mathrm{e}^{-|y|},\ y \in \mathbb{R}$$
$$f_{Y|X}(y \mid 1) = \frac{1}{4}\mathrm{e}^{-\frac{|y-2|}{2}},\ y \in \mathbb{R}.$$

Construct a decision rule $r : \mathbb{R} \to \{0, 1\}$ that minimizes $\mathbb{P}(r(Y) \neq X) = \mathbb{E}[I\{r(Y) \neq X\}]$ (use the MAP rule).

(b) [10] We consider a standard hypothesis testing problem in which we have

$$Y \mid X = 0 = \begin{cases} a, & \text{w.p. } \frac{1}{4} \\ b, & \text{w.p. } \frac{1}{4} \\ c, & \text{w.p. } \frac{1}{2}, \end{cases} \qquad \text{and} \qquad Y \mid X = 1 = \begin{cases} a, & \text{w.p. } \frac{1}{2} \\ b, & \text{w.p. } \frac{1}{4} \\ c, & \text{w.p. } \frac{1}{4}. \end{cases}$$

Design a randomized decision rule $r : \{a, b, c\} \to \{0, 1\}$ in order to minimize $\mathbb{P}(r(Y) = 0 \mid X = 1)$ subject to $\mathbb{P}(r(Y) = 1 \mid X = 0) \leq 0.3$.

Problem 4: Estimating the Slope of a Line [15]

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d., where $X_1 \sim \mathcal{N}(0,1)$ and $Y_1 \mid X_1 = x \sim \mathcal{N}(x\theta, 1)$ where $\theta \in \mathbb{R}$ is unknown.

(a) **[8]** Compute the MLE of θ given the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$.

(b) **[7]** Compute $\mathbb{E}[X_1 \mid Y_1]$.

Problem 5: Estimating Jointly Gaussian Random Variables [25]

Consider the Gaussian random vector

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right).$$

(a) [**4**] Draw on the plane the random variables as vectors in the Hilbert space of random variables. Make sure to note the length of the vectors, as well as the angle that they form.

(b) [**4**] Find $\mathbb{E}[X \mid Y]$.

(c) [**5**] Create a random variable Z which is a function of X and Y such that

- Z and Y are independent.
- The vector $\begin{bmatrix} Z \\ Y \end{bmatrix}$ has exactly the same information content as the vector $\begin{bmatrix} X \\ Y \end{bmatrix}$, in the sense that you can recover the latter vector from the former.

Show that your Z satisfies these properties.

(d) **[6]** Let $U = Y^3$. Find $\mathbb{E}[X \mid U]$.

Hint: This requires almost no computation, but make sure you justify your answer.

(e) **[6]** Let $V = Y^2$. Find $\mathbb{E}[X \mid V]$.

Hint: This requires almost no computation as well, but first understand why it is different from (c).

Problem 6: Recursive Parameter Estimation [10]

Consider the stochastic dynamical system

$$X_{n+1} = X_n,$$
$$Y_n = X_n + W_n,$$

for each $n \in \mathbb{N}$, where $X_0, W_0, W_1, W_2, \ldots$ are i.i.d. $\mathcal{N}(0, 1)$. At time $n - 1$, suppose that we have computed the state estimate $\hat{X}_{n-1|n-1}$. Then, at time n, we see the observation Y_n.

(a) [3] How do we compute the innovation at time n?

(b) [7] We know that the Kalman filter equations recursively calculate $\hat{X}_{n|n}$ in "real time". Suppose we are willing to tolerate a delay of one sample in our estimate, i.e. we can wait until *after* observing Y_{n+1} to estimate X_n for $n = 1, 2, 3, \ldots$. Let $\hat{X}_{n|n}$ be the MMSE of X_n at time n (i.e. based on $Y^{(n)} = Y_0, Y_1, \ldots, Y_n$), and let $\sigma^2_{n|n}$ be our estimation error at time n (i.e. $\sigma^2_{n|n} = \mathbb{E}[(X_n - \hat{X}_{n|n})^2]$). Consider the update equation

$$\hat{X}_{n|n+1} = \hat{X}_{n|n} + a(Y_{n+1} - \hat{X}_{n|n}).$$

What is the optimal value of a to minimize $\sigma^2_{n|n+1} = \mathbb{E}[(X_n - \hat{X}_{n|n+1})^2]$?

Problem 7: HMM Estimation [15]

Consider a HMM $(X_n)_{n \in \mathbb{N}}$ with state space $\{0, 1\}$ and transitions $P(0, 1) = P(1, 0) = p \in [0, 1]$. The hidden state is observed through a Binary Symmetric Channel (BSC) with error probability 1/3. Assume that the initial state is equally likely to be 0 or 1. The sequence of observations is $(1, 1, 1)$.

(a) [5] Suppose that it is known that $p = 3/4$. What is the most likely sequence of hidden states?

(b) [5] Now, suppose that p is unknown. We will try applying the hard EM algorithm to this problem. We start by initializing $\hat{p}^{(0)} = 3/4$. In the E step, we fill in the most likely values of the hidden variables (X_0, X_1, X_2) given the observations, assuming that the parameter is $\hat{p}^{(0)}$. Carry out the E step.

(c) [5] In the M step, we compute the MLE of p assuming the filled-in values for the hidden variables computed in the E step. Carry out the M step.