EECS 126 Probability and Random Processes — UC Berkeley, Spring 2021
Thomas Courtade — May 10, 2021

Final

Last Name	First Name	SID

Rules.

- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- You have 160 minutes to complete the exam and 10 minutes exclusively for submitting your exam to Gradescope. (DSP students with $X\%$ time accomodation should spend $160 \cdot X\%$ time on the exam and 10 minutes to submit).
- Collaboration with others is strictly prohibited.
- You should not discuss the exam with anyone (this includes your roommate, your parents, social media, reddit, etc.) until 24 hours after the exam concludes (May 11, 2:30pm).
- You may reference your notes, the textbook, and any material that can be found through the course website. You may use Google to search for general knowledge or use calculators. However, searching for a question is not allowed.
- For any clarifications you have, please create a private Piazza post. We will have a Google Doc that shows our official clarifications.

Problem	points earned	out of
Honor Code		5
Problem 1		10
Problem 2		10
Problem 3		10
Problem 4		13
Problem 5		14
Problem 6		12
Problem 7		11
Problem 8		15
Total		100

Honor Code [5 points]

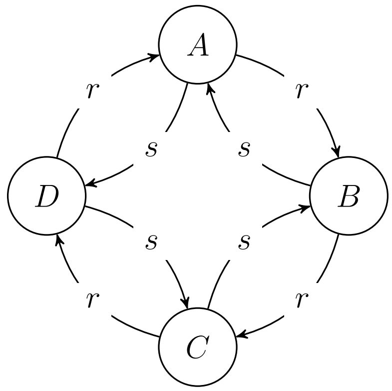
Please copy the following word for word, and sign afterwards.

By my honor, I confirm that

1. *this work is my own original work;*
2. *I have not and will not discuss this exam with anyone during the exam and for 24 hours after the exam;*
3. *I have not and will not Google/search for any of these exam problems.*

1 Going in Circles [4 + 6]

Consider the continuous time Markov Chain illustrated below, where r and s denote transition rates.



a) Assume that you start in state A. What is the probability that you visit each state exactly once and then return to A?

We need to either go $A \to B \to C \to D \to A$ in the jump chain, or $A \to D \to C \to B \to A$ in the jump chain. For the jump chain, the probability of transitioning in the clockwise direction is $r/(r+s)$ for each state, and similarly $s/(r+s)$ to move counterclockwise. Hence, the probability is

$$\left(\frac{r}{r+s}\right)^4 + \left(\frac{s}{r+s}\right)^4$$

b) Let $s = 1$ and $r = 2$. Starting at state $X_0 = A$, compute the expected time T it takes for the chain to enter state C.

The first-step equations are:

$$\begin{aligned}
\beta(A) &= \frac{1}{3} + \frac{2}{3}\beta(B) + \frac{1}{3}\beta(D) \\
\beta(B) &= \frac{1}{3} + \frac{1}{3}\beta(A) \\
\beta(D) &= \frac{1}{3} + \frac{2}{3}\beta(A),
\end{aligned}$$

where $\beta(i) := \mathbb{E}[T|X_0 = i]$. Substituting the latter two equations into the first gives:

$$\begin{aligned}
\beta(A) &= \frac{1}{3} + \frac{2}{3}(\frac{1}{3} + \frac{1}{3}\beta(A)) + \frac{1}{3}(\frac{1}{3} + \frac{2}{3}\beta(A)) \\
\beta(A) &= \frac{1}{3} + \frac{2}{9} + \frac{2}{9}\beta(A) + \frac{1}{9} + \frac{2}{9}\beta(A) \\
\frac{5}{9}\beta(A) &= \frac{6}{9} \\
\beta(A) &= \frac{6}{5}
\end{aligned}$$

2 Ants [5 + 5]

The problems below are unrelated, except that they both involve ants on a cube. By this, we mean the 3-d cube with vertices $\{000, 001, 010, \ldots, 111\}$, where vertices are connected by an edge if their labels differ in exactly one position.

a) Assume an ant starts at vertex 000, and takes a random walk on the cube. That is, the next vertex is chosen randomly from all neighbors of the current vertex. If the ant collects $G_i \sim \text{Binomial}(n, p)$ grains of sugar on the i-th edge it traverses, what is the expected total number of sugar grains collected by the time it returns to vertex 000? Assume the G_i's are independent of the number of edges traversed.

b) Assume the ant again takes a random walk on the cube, starting at 000. However, this time, assume it takes the ant $T_i \sim \text{Exp}(1/2)$ to traverse the i-th edge, and it starts traversing the next edge as soon as it completes the previous. At a given time $t \geq 0$, what is the distribution of the total number of traversals completed (exclude the one in progress)?

a) Let N be the number of edges traversed before returning to the starting state. The Big Theorem for MCs gives $\mathbb{E}[N] = 8$, since the stationary distribution is uniform over states. Hence, by iterated expectation:

$$\mathbb{E}[\sum_{i=1}^{N} G_i] = \mathbb{E}[\sum_{i=1}^{N} \mathbb{E}[G_i|N]] = \mathbb{E}[Nnp] = 8np.$$

b) Since the time to cross edges are exponential random variables, the number of edges traversals completed forms a Poisson process with rate $1/2$. Hence, the number of edges traversed up to time t has distribution Poisson$(t/2)$.

3 Minimum Mean-Square Estimation [5 + 5]

Let X_n represent the position of a parked car along a long road at time instant $n \geq 1$, and let Y_n represent the GPS reading of the the car's position. We assume $X_0 \sim N(0,1)$ and updates according to the following equations:

$$\begin{aligned} X_n &= X_{n-1} + V_n \\ Y_n &= X_n + W_n, \qquad n \geq 1, \end{aligned}$$

where $V_n \sim N(0,1), W_n \sim N(0,1)$ are independent sources of noise.

a) Suppose we somehow observe V_3 and only V_3. Compute the minimum mean-square error estimate (MMSE) of X_4 given V_3.

b) Now, suppose we observe Y_3 and only Y_3. Compute the MMSE of X_4 given Y_3.

a) By linearity of expectation, all X_n's are zero-mean. Making substitutions using definitions, the MMSE is

$$\mathbb{E}[X_4|V_3] = \mathbb{E}[X_3 + V_4|V_3] = \mathbb{E}[X_2 + V_3 + V_4|V_3] = V_3.$$

The last step follows by linearity of expectation and independence.

b) Note that X_3 and Y_3 are jointly gaussian random variables since the sum of gaussians is still gaussian. Thus, $MMSE = LLSE$ and

$$\begin{aligned} L[X_4|Y_3] &= E[X_3] + \frac{Cov(X_4, Y_3)}{Var(Y_3)}(Y_3 - E[Y_3]) \\ &= E[X_3] + \frac{Cov(X_3 + V_4, X_3 + W_3)}{Var(Y_3)}(Y_3 - E[Y_3]) \\ &= \frac{Var(X_3)}{Var(X_3) + Var(W_3)} Y_3 \\ &= \frac{4}{5} Y_3 \end{aligned}$$

The last line follows since $Var(W_3) = 1$ and

$$Var(X_3) = Var(X_2 + V_3) = Var(X_1 + V_2 + V_3) = Var(X_0 + V_1 + V_2 + V_3) = 4$$

where the last equality follows from independence and the variances given in the problem setup.

4 Study Habits [2 + 4 + 7]

Kevin's system for working on his weekly 126 homework can be modeled as follows. Let X_i be the indicator that Kevin is working on his 126 homework on night $i \in \{1, 2, \dots, 7\}$, where homework is due at midnight on night $i = 7$. The three nights before the weekly homework deadline, the joint distribution is

$$(X_5, X_6, X_7) = \begin{cases} (0,0,0) & \text{w.p. } 1/4 \\ (1,1,0) & \text{w.p. } 1/4 \\ (0,1,1) & \text{w.p. } 1/4 \\ (1,0,1) & \text{w.p. } 1/4, \end{cases}$$

Otherwise (on the other 4 nights), he works on the homework independently with probability 1/3.

a) Are the X_i's independent?

b) The semester lasts for 15 weeks, and Kevin implements the same system each week. What is the expected number of nights that Kevin spends working on homework?

c) Michael hits up Kevin to hang out 2 nights in a row, but is dismayed to find his friend working on the 126 problem set both nights. What is the probability that both nights Michael contacts Kevin are within 3 days of the homework deadline? (i.e., Michael first contacts Kevin on either night 5 or night 6.)

[Assume Michael first contacts Kevin on a night uniformly selected from $\{1, 2, \dots, 7\}$. Moreover, assume that there is a homework every week, so if Michael first hits up Kevin on night $i = 7$, he observes (X_7, X_1).]

a) No. For example, the event $\{X_5 = X_6 = X_7 = 1\}$ has zero probability under the model, but would have positive probability if the random variables were independent.

b) Note that

$$\mathbb{E}[X_i] = \begin{cases} 1/3 & \text{if } 1 \le i \le 4 \\ 1/2 & \text{if } 5 \le i \le 7. \end{cases}$$

So, by linearity of expectation, the expected number of nights worked in a given week is

$$\mathbb{E}[\sum_{i=1}^{7} X_i] = 4/3 + 3/2 = 17/6.$$

By linearity of expectation again, the number of nights worked in a semester is:

$$15 \times 17/6 = 42.5 \text{ nights}.$$

c) Let A be the event that both nights are within three days of the deadline, B be the event that one is, and C be the event that both are not. Let E be the observation that Kevin is studying 2 nights in a row. By Bayes rule, we have

$$\begin{aligned}
\Pr[A|E] &= \frac{\Pr[E|A]\Pr[A]}{\Pr[E|A]\Pr[A] + \Pr[E|B]\Pr[B] + \Pr[E|C]\Pr[C]} \\
&= \frac{1/4 \cdot 2/7}{1/4 \cdot 2/7 + 1/3 \cdot 1/2 \cdot 2/7 + 1/3 \cdot 1/3 \cdot 3/7} \\
&= \frac{1/2}{1/2 + 1/3 + 1/3} = 3/7.
\end{aligned}$$

5 Gaussian Distances [5 + 5 + 4]

a) Let $Z \sim \mathcal{N}(0,1)$ be a standard normal random variable. For $\lambda > 0$, show that

$$P(Z > \lambda) \leq e^{-\lambda^2/2}.$$

Hint: The m.g.f. of a standard normal is given by $M_Z(t) = e^{t^2/2}$.

b) Let $X = (X_1, X_2, \ldots, X_n) \sim N(0, I_n)$, where I_n denotes the $n \times n$ identity matrix. *Without appealing to expressions for probability densities*, show that the distribution of X is rotation invariant. More specifically, show that if U is an orthogonal matrix (i.e., $UU^T = U^T U = I_n$) then UX has the same distribution as X.

c) Let X be as above. Suppose $\mathcal{V} \subset \mathbb{R}^n$ is subspace of dimension $n-1$. We let $d(x, \mathcal{V}) := \min_{y \in \mathcal{V}} |x - y|$ denote the Euclidean distance between a vector $x \in \mathbb{R}^n$ and the subspace $\mathcal{V}$. Show that

$$P(d(X, \mathcal{V}) > \lambda) \leq 2e^{-\lambda^2/2}.$$

[You can consider the case $n = 2$ for full credit if you aren't comfortable with the linear algebra in n dimensions.]

a) We use a Chernoff bound:

$$P(Z > \lambda) = P(e^{tZ} > e^{t\lambda}) \leq \frac{\mathbb{E}[e^{tZ}]}{e^{t\lambda}} = e^{t^2/2 - t\lambda}.$$

Now, since $\exp(\cdot)$ is monotonic, it suffices to optimize the exponent; setting the derivative to zero gives us

$$t - \lambda = 0 \implies t = \lambda.$$

Plugging this in, we get

$$P(Z > \lambda) \leq e^{\lambda^2/2 - \lambda^2} = e^{-\lambda^2/2}.$$

b) Since UX is a linear transformation of iid gaussians, it is itself a gaussian vector. Gaussian vectors have distributions parameterized completely by mean and covariance. By linearity of expectation,

$$\mathbb{E}[UX] = U\,\mathbb{E}[X] = 0,$$

so UX has mean zero. Again by linearity of expectation, its covariance is

$$\mathbb{E}[(UX)(UX)^T] = \mathbb{E}[UXX^TU^T] = U\,\mathbb{E}[XX^T]U^T = UU^T = I_n.$$

Hence, $UX \sim N(0, I_n)$ as claimed.

c) We just showed in the previous part that the distribution of X is rotationally invariant, so without loss of generality, we can rotate the coordinate space (an orthogonal change of basis) and assume the vector $(1, 0, 0, \ldots, 0)$ is normal to $\mathcal{V}$. Hence, $d(X, \mathcal{V}) = |X_1|$, and using part (a) we have

$$P(|d(X, \mathcal{V})| > \lambda) = P(X_n > \lambda) + P(X_n < -\lambda) \leq 2e^{-\lambda^2/2}.$$

[If you aren't comfortable with the concept of change of basis, the idea is clear by drawing a picture in dimension 2, where $\mathcal{V}$ is a line passing through the origin, and you just rotate the picture until that line is in the vertical direction. None of this changes the distribution of X because it is rotationally invariant, as shown in part (b).]

6 Bot or Not [8 + 4]

Reina is playing chess online and trying to figure out what type of bot she is playing against. For the next move, a naïve bot will select which type of piece to play (king, rook, bishop, queen, knight, and pawn) according to a uniform distribution. An advanced bot will select the type of piece based on the following distribution:

$$\text{Y} = \begin{cases} \text{Rook} & \text{w.p. } 0.07 \\ \text{Queen} & \text{w.p. } 0.18 \\ \text{Knight} & \text{w.p. } 0.2 \\ \text{Pawn} & \text{w.p. } 0.3 \\ \text{Bishop} & \text{w.p. } 0.15 \\ \text{King} & \text{w.p. } 0.1 \end{cases}$$

Assume the null hypothesis corresponds to playing against a naïve bot, and you'll accept or reject this hypothesis after seeing which piece the bot plays.

a) Construct a decision rule that minimizes probability of false negative (i.e., Type II error rate) subject to a false positive rate (i.e., Type I error rate) of at most $1/4$.

Note: A decision rule in this context is a (possibly randomized) mapping from pieces to bot-type.

b) Suppose now that you are given the prior ($\pi_0 = 1/3, \pi_1 = 2/3$), where π_0 is the prior probability of the null hypothesis being correct. Under your decision rule, what is the probability of incorrectly determining the bot type?

a) The likelihood ratio is given by

$$L(Y) = \begin{cases} 6 \times 0.07 & \text{if } Y = \text{Rook} \\ 6 \times 0.18 & \text{if } Y = \text{Queen} \\ 6 \times 0.2 & \text{if } Y = \text{Knight} \\ 6 \times 0.3 & \text{if } Y = \text{Pawn} \\ 6 \times 0.15 & \text{if } Y = \text{Bishop} \\ 6 \times 0.1 & \text{if } Y = \text{King} \end{cases}$$

So, any deterministic threshold test will simply declare the bot to be advanced if the piece is among the k most likely played by an advanced bot (k is determined by our choice of threshold). In this case, the probability of false alarm is precisely $k/6$. To get PFA$= 1/4 = 1/2 \times 2/6 + 1/2 \times 1/6$, we should randomize between the tests selecting the $k = 2$ and $k = 1$ most likely pieces, where we choose each test with probability $1/2$. Thus, our decision rule is as follows:

- If we see a Pawn, declare advanced bot.
- If we see a Knight, declare a bot type with probability $1/2$ each.

- If we see any other piece, declare a naive bot.

b) Let r denote our decision rule above, and let $X \in \{0, 1\}$ denote the correct hypothesis. For the given prior, our probability of being incorrect is

$$\begin{aligned}
\Pr[r(Y) \neq X] &= \pi_0 \Pr[r(Y) = 1 | X = 0] + \pi_1 \Pr[r(Y) = 0 | X = 1] \\
&= \frac{1}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot (1 - (0.3 + 0.2/2)) \\
&= \frac{1}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{6}{10} \approx 0.483.
\end{aligned}$$

7 Random(?) Graphs [2 + 9]

Consider a graph $G = (V, E)$ on n vertices. The vertex set V is deterministically partitioned as $V = V_1 \cup V_2$, where V_1 and V_2 are disjoint and of unknown sizes, but we assume $|V_i| \geq n/100$ for $i = 1, 2$.

The edges of G are placed randomly as follows. An edge (u, v) appears with probability $p < 1$ if vertices u and v are both in V_1 or both in V_2. Otherwise, an edge is placed with probability one. Assume all edges are placed independently, and that p does not depend on n.

a) Is G an Erdös–Rényi random graph?

b) Suppose you don't know which vertices are in V_1 or V_2. You try to determine this by finding a partition $\hat{V}_1 \cup \hat{V}_2 = V$ such that each $\hat{V}_i$ has at least one vertex, and $(u, v) \in E$ for all $u \in \hat{V}_1, v \in \hat{V}_2$. If there are multiple candidate partitions satisfying this criteria, you choose one arbitrarily. Your choice is considered "correct" if $\hat{V}_1 = V_1$ or $\hat{V}_1 = V_2$, since both identify the same partition of V.

   In the limit as $n \to \infty$, what is the probability that the above procedure recovers the correct partition? Answers without complete justification will be penalized.

   (*Hint:* Consider the complement G' of the graph G, where an edge exists in G' iff it does not exist in the original graph G.)

a) No; edge placements in G are not identically distributed.

b) If we consider the complement graph G', there will be no edges crossing from V_1 to V_2. Since $p < 1$, edges appear in the compliment graphs on V_1 and V_2 with probability $q := 1 - p > 0$, in an iid fashion. Hence, the picture we have is one of two disjoint Erdos-Renyi graphs $G_1 \sim G(|V_1|, q)$ and $G_2 \sim G(|V_2|, q)$. For n large enough, $q \gg \frac{\log |V_i|}{|V_i|}$, so each G_i will be connected with probability approaching 1. In this case, there is no ambiguity about which vertices belong to which G_i (in other words, G' consists of two disjoint connected components). So, the proposed procedure will identify the correct partition with probability approaching 1 as $n \to \infty$.

8 Counting Birds [6 + 5 + 4]

Han wants to estimate N, the total number of birds in a nearby park. On her first visit to the park, Han observes 17 birds, and marks each of them with a dab of paint. On her second visit to the park, Han observes 20 birds, and 6 of them carry the mark (assume no marks disappear).

On each visit, we assume that each of the N birds is observed independently with probability $1/2$.

a) As a function of N, what is the likelihood of Han's observations?

b) What is the maximum-likelihood estimate of N given Han's observations?

[Hint: Since N is discrete, we can't differentiate and set the derivative equal to zero. Instead, consider the ratio:

$$\frac{\Pr[\text{observation}|N]}{\Pr[\text{observation}|N+1]}$$

and find the smallest N for which this ratio is greater than 1.]

c) Assume N has prior distribution $N \sim \text{Uniform}\{50, 51, \ldots, 100\}$. What is the MAP estimate of N?

a) According to the model, the likelihood of Han's observations are

$$\Pr[\text{observations}|N] = \binom{N}{17} 2^{-N} \binom{N-17}{14} \binom{17}{6} 2^{-N} = \frac{N!}{(N-31)!} 2^{-2N} \frac{1}{6! \times 11! \times 14!}.$$

This is because: (i) there are $\binom{N}{17}$ configurations of birds she could have seen on the first day, each appearing with probability 2^{-N}; (ii) There are $\binom{N-17}{14}\binom{17}{6}$ configurations of birds she could have seen on the second day, each appearing with probability 2^{-N}.

b) Consider the ratio

$$\frac{\Pr[\text{observations}|N]}{\Pr[\text{observations}|N+1]} = 4 \times \frac{N-30}{N+1}.$$

This ratio being greater than 1 is equivalent to $N > \frac{121}{3} \approx 40.33$, so our likelihoods reach a maximum at $N = 41$, which is the MLE.

c) Since our prior is uniform, MLE = MAP. But here $N = 41$ has probability zero, so we should pick the value of N that maximizes likelihood, while having positive probability. So we pick $N = 50$.