EECS 126 Probability and Random Processes — University of California, Berkeley: Fall 2018

Kannan Ramchandran and Abhay Parekh — December 12, 2018

Final

Last Name	First Name	SID

- You have 5 minutes to read the exam and 175 minutes to complete this exam.
- The maximum you can score is 134, 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		45
Problem 2		20
Problem 3		10
Problem 4		12
Problem 5		7
Problem 6		20
Problem 7		20
Total		100 (+34)

Problem 1: Answer these questions briefly but clearly. [45]

1. **Maximum Variance [5]**
   Let X be a random variable that takes value between 0 and c, where c is positive-valued (i.e. $\mathbb{P}(0 \leq X \leq c) = 1$). What is the maximum value of the variance of X? Provide an example which achieves this bound. You do not need to prove that your bound (if correct) is tight.

2. **Min and Max of Uniform Distribution [5]** Let X and Y be independent random variables distributed as Uniform$[0, 1]$. Let $U = \min\{X, Y\}$ and $V = \max\{X, Y\}$. Find $\text{cov}(U, V)$.

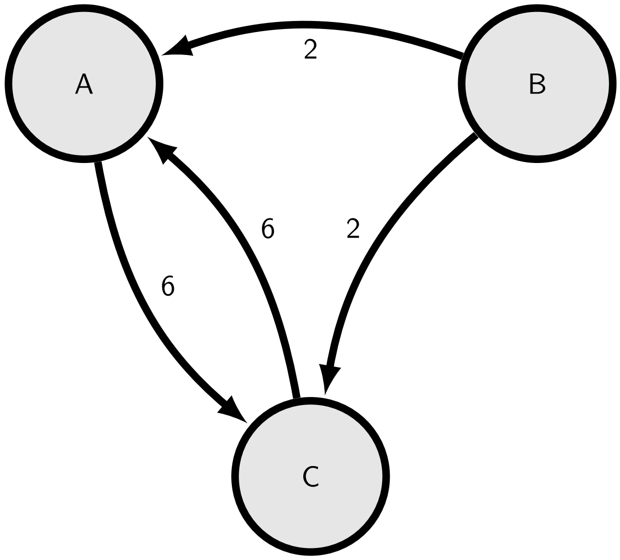
3. **Correlation Coefficients [5]**
   Let X, Y, Z be jointly Gaussian zero–mean random variables such that X is conditionally independent of Z given Y. Given that the correlation coefficients of (X, Y) and (Y, Z) are ρ_1 and ρ_2, find the correlation coefficient of (X, Z). *Hint: The answer is in a fairly simple form; use the law of iterated expectation.*

4. **Short Questions (Justify, no points for only answer.) [6]**

(a) True or False? For Zero Mean Jointly Gaussian RVs X, Y and Z, if $L(X|Y, Z) = L(X|Y) + L(X|Z)$, then Y and Z are independent RVs.

(b) If X is a Poisson Process of rate λ and has N arrivals in $(0, T)$, what is the joint distribution of the first N arrival times?

5. **CTMC [7]**
Consider the CTMC shown below. Write out the transition matrix, find the stationary distribution and then find the corresponding DTMC which has the same stationary distribution as this chain.



6. **Deterministic Poisson Splitting [5]**

Customers arrive at a store in a Poisson process, $N(t), (t > 0)$, with rate λ. There are two queues, Q_1 and Q_2. Instead of random assignment to the queues, the first customer is deterministically assigned to Q_1, the next is assigned to Q_2, and so on; that is, the customers are assigned alternately to the two queues. Are the arrival processes to the individual queues Poisson? (If yes, provide the rate of the process. If no, show why not.)

7. **Petersburg Revisited [7]**

Recall the St. Petersburg "paradox" example from lecture. Formally, let X be a random variable representing the payoff from a random game such that for $i = 1, ..., P(X = 2^i) = \frac{1}{2^i}$. In lecture, we showed that $\mathbb{E}[X]$ is infinite, but this does not seem to be a reasonable way to model the "fair price" of the game. Here, we explore a different approach. Now, let X_k be i.i.d. realizations of this game at time step k. At each time step, according to some fixed $c \in \mathbb{R}$, define

$$S_n = \prod_{i=1}^{n} \frac{X_i}{c}$$

(a) What is the distribution of $\log_2 X_i$? (Specify parameters, if any. No justification needed.)

(b) Show that $\mathbb{E}[\log_2(X_i)] = 2, \forall i$. If needed, use without proof that $\sum_{i=1}^{\infty} \frac{i}{2^i} = 2$.

(c) Show that $\lim_{n\to\infty} \log_2(S_n)$ is either $-\infty$ or ∞, w.p. 1 according to if $c < c^*$ or $c > c^*$ for some fixed c^*. What is this value c^*?

8. **Independent Sum Entropy [5]**

Let $X_1, ..., X_n$ be independent random variables. Let $Y = \sum_{i=1}^{n} X_i$.

(a) Argue that $\forall i$, $H(Y) \geq H(X_i)$.

(b) Give an example where $H(Y) = \sum_{i=1}^{n} H(X_i)$.

Problem 2: Random Graphs and Markov Chains [20]

Assume $N \geq 3$ is a fixed positive integer and G_0 is a graph on N vertices $\{1, 2, \ldots, N\}$ with no edges (empty graph). At each time step $n \geq 1$, starting from the graph G_{n-1}, we pick a pair (i, j), $1 \leq i < j \leq N$ uniformly at random among the $\binom{N}{2}$ such pairs. Then, with probability $1-p$, we do nothing, and with probability p, we alter the edge between vertices i and j; that is, if there is an edge between i and j, we remove it and if there is not an edge, we place an edge. Here, $p \in (0, 1)$ is fixed. Let G_n be the resulting graph. We continue this process inductively, i.e. we generate G_1 from G_0, then G_2 from G_1, and so on.

1. Argue that $(G_n : n \geq 0)$ is a Markov Chain. What is the state space? What is the size of this state space? What are the transition probabilities?

2. Classify this chain in terms of its periodicity, reducibility, and whether it is transient, positive recurrent or null recurrent (Justify your answers for full credit).

3. What are the stationary distribution(s) of this Markov Chain? [Hint: Show that the Erdős–Rényi distribution $\mathcal{G}(N, q)$ with an appropriate value of q is the stationary distribution.]

4. Assume that $N = 3$. Let T be the first time such that G_T is the complete graph (the graph on 3 vertices with all the 3 possible edges present). Find $\mathbb{E}(T)$.

Problem 3: MAP with Gaussians [10]

A disease has 2 strains, 0 and 1, which occur with prior probability p_0 and $p_1 = 1 - p_0$ respectively. *For both parts of this problem, you are allowed to leave your answer in terms of* $\Phi(x)$*, the CDF of the standard normal distribution.*

1. A noisy test is developed to find which strain is present for patients with the disease. Let $X \in \{0, 1\}$ be the random variable which denotes the strain. The output of the test is a random variable Y_1, such that $Y_1 = 5 - 4X + Z_1$, where $Z_1 \sim \mathcal{N}(0, \sigma^2)$ and is independent of the strain X. Give a MAP decision rule to output $\hat{X}$, your best guess for X, given Y_1, and compute $\mathbb{P}(\hat{X} \neq 0 | X = 0)$.

2. A medical researcher proposes a new measurement procedure: he observes Y_1 as done previously, and in addition, "creates" a new measurement, $Y_2 = Y_1 + Z_2$. Assume $Z_2 \sim \mathcal{N}(0, \sigma^2)$ is independent of X and Z_1. Now, find the MAP rule in terms of the joint observation (y_1, y_2) and compute $\mathbb{P}(\hat{X} \neq 0 | X = 0)$.

Problem 4: Hypothesis Testing with Gaussians [12]

For this problem also, you may leave your answer in terms of the Guassian CDF $\Phi(x)$.

1. We are told that a random variable X is either $\mathcal{N}(0,1)$ (null hypothesis) or $\mathcal{N}(10,1)$ (alternate hypothesis). We want the probability of false alarm to be no more than 2.5%. What is the Neyman-Pearson optimal test? What is the probability of correct detection for this threshold?

2. Now, we are told that a random variable Y is either $\mathcal{N}(0,1)$ (null hypothesis) or $\mathcal{N}(0,2)$ (alternate hypothesis). We want the probability of false alarm to be no more than 5%. What is the Neyman-Pearson optimal test? What is the probability of correct detection for this threshold?

Problem 5: Gaussian Product CLT [7]

Let $X_1, ..., X_n \overset{i.i.d.}{\sim} \text{Lognormal}(\mu, \sigma)$. Let $Y_k := (\Pi_{i=1}^{k} X_i)^{1/k}$. *Recall: if* X *is log-normally distributed, then* $\ln(X)$ *is* $\mathcal{N}(\mu, \sigma)$.

1. Find $\mathbb{E}[\ln(Y_k)]$.

2. Find a lower bound on n such that $\mathbb{P}(|\ln(Y_n) - \mathbb{E}[\ln(Y_n)]| > 0.01) < 0.05$. You may leave your answer in terms of $\Phi(x)$, the normal CDF.

Problem 6: LLSE and Kalman Filter [20]

Consider a sensor network comprising n sensors that take noisy measurements of a temperature variable X as follows: $Y_i = X + W_i$, where $X \sim \mathcal{N}(0, 10)$ and W_i's are i.i.d. $\mathcal{N}(0, 1)$ that model the noise in the system.

1. Let $\hat{X}_{LLSE} = \alpha_1 Y_1 + \alpha_2 Y_2 + \ldots + \alpha_n Y_n$. Find α_i for $i = \{1, \ldots, n\}$. *(Hint: Do it for $n = 2$ first and then generalize).*

2. Suppose $n = 2$. I want to form $L[X|Y_1, Y_2]$ in an online fashion by first considering Y_1 and then Y_2 as follows:

$$L[X|Y_1, Y_2] = L[X|Y_1] + L[X|\tilde{Y_2}].$$

   What is $\tilde{Y_2}$? Draw a geometric picture relating Y_1, Y_2 and $\tilde{Y_2}$.

3. Now I want to estimate X recursively by taking the measurements $Y_1, Y_2, \ldots, Y_n$ in an online fashion using a Kalman Filter based approach. Note that the state-space equations degenerate to:

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

   We will use the usual notation seen in lecture. $\hat{X}_{n|n}$ is the best estimate of X_n given $Y_1, Y_2, \ldots, Y_n$. $\hat{X}_{n|n-1}$ is the best estimate of X_n given $Y_1, Y_2, ...Y_{n-1}$ and $\sigma^2_{n|n} = E((X_n - \hat{X}_{n|n})^2)$, etc.

Suppose I initialize $\hat{X}_{1|0} = 0$ and $\sigma^2_{1|0} = 10$ (i.e, variance of X) in the Kalman equations:

$$\hat{X}_{n|n} = \hat{X}_{n|n-1} + k_n(Y_n - \hat{X}_{n|n-1})$$

$$k_n = \frac{\sigma^2_{n|n-1}}{\sigma^2_{n|n-1} + \sigma^2_w}$$

$$\sigma^2_{n|n} = \sigma^2_{n|n-1}(1 - k_n)$$

(a) What are $\hat{X}_{1|1}, \sigma^2_{1|1}, \hat{X}_{2|2}$, and $\sigma^2_{2|2}$?

(b) In the limit as $n \to \infty$, what are k_n and $\sigma^2_{n|n}$?

Problem 7: HMMs and EM [20]

1. There are two identical-looking coins A and B whose biases (probability of Heads) are $\theta_A = 0.4$ and $\theta_B = 0.8$ respectively. Let X_k be the coin at time step k. A Markov Chain with transition probabilities given below describes the coin-picking process: $P(X_{k+1} = A|X_k = B) = 0.2$, $P(X_{k+1} = B|X_k = A) = 0.3$ for $k = \{0, 1 \ldots, \}$. Now, let the initial state X_0 be A. At each time step, we observe the result of flipping the current coin (without knowing which coin it was). The observed sequence of tosses is H,T,T.

   (a) What is the most likely sequence of coin labels picked?

   (b) What is the most likely coin label corresponding to the *second* toss? Is it consistent with the answer in (a)? Does it need to be? Explain.

2. Now suppose that you do not know the true biases of the two coins and want to estimate them. At each time step, you pick one of the two coins equally at random and toss it once and observe whether it is Heads or Tails. You then replace the coin and repeat the experiment 5 times. Suppose you observe H, T, T, H, H.

   (a) Using the Hard EM algorithm with initial guess $\theta_A = 0.4, \theta_B = 0.8$, what will be your converged estimates of the biases of the coins?

   (b) Now you use the Soft EM algorithm with the same initial guesses. What will be the estimates for θ_A, θ_B after one iteration?