EECS 126 Probability and Random Processes — University of California, Berkeley: Fall 2017
Abhay Parekh and Jean Walrand — December 14, 2017

Final Exam

Last Name	First Name	SID

Rules.

- You have 170 minutes (3:10pm - 6:00pm) to complete this exam.
- The maximum you can score is 130.
- The exam is not open book, but you are allowed one side of a sheet of handwritten notes; calculators will be allowed. No phones.
- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.

Please read the following remarks carefully.

- 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		40
Problem 2		30
Problem 3		20
Problem 4		20
Problem 5		20
Problem 6		5 Bonus
Total		130 (+5 Bonus)

Problem 1 [40] Short Answers

(a) [5] Alice and Bob play the following game. Alice tosses n identical coins, each of which comes up Heads with probability p. All coins that come up Tails are removed and Alice flips the remaining coins again. If there are X coins that come up Heads this time, Bob gets X dollars. Find the pmf for X. Your answer should contain no summation signs.

A given coin makes it through two tosses with probability p^2. Thus

$$p_X(k) = \binom{n}{k} p^{2k}(1-p^2)^{n-k}$$

(b) [5] The joint distribution of two random variables X and Y is given by: $p_{X,Y}(0,0) = p_{X,Y}(1,0) = 0.2$, $p_{X,Y}(0,1) = p_{X,Y}(1,1) = 0.3$. What is $E[\frac{10}{p_{Y|X}(Y|X)}]$?

$$\begin{aligned} E[\frac{10}{p_{Y|X}(y|x)}] &= \sum_{x=0}^{1}\sum_{y=0}^{1} p_{X,Y}(x,y)\frac{10}{p_{X,Y}(x,y)/p_X(x)} \\ &= \sum_{x=0}^{1}\sum_{y=0}^{1} 10 p_X(x) = \sum_{y=0}^{1} 10 = 20. \end{aligned}$$

(c) [5] The Figure describes $F_X(x)$, the CDF of X. Find $E(X^2)$.

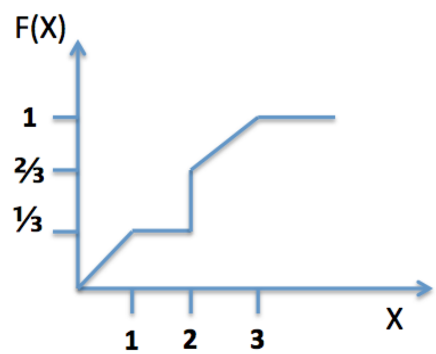

Figure 1: CDF

X is a mixed random variable. It is 2 with probability $1/3$ and in $(1,2)$ with probability 0. It is equally likely (with probability $1/3$) to be in $[0,1]$ and $(2,3]$ and is uniformly distributed in both intervals. Thus

$$E[X^2] = \frac{4}{3} + \int_0^1 x^2 \frac{1}{3} dx + \int_2^3 x^2 \frac{1}{3} dx = \frac{4}{3} + \frac{1}{9} + \frac{19}{9} = \frac{32}{9}.$$

(d) **[5]** Email to Bob gets automatically classified and those considered to be Spam are routed to a Spam folder. Email not classified as spam is routed to an Inbox. Regular email arrives as a Poisson Process with rate 10 msgs/hour and Spam arrives at a rate of 20 msgs/hour. A regular email is classified as spam with probability 0.05 and a spam email is classified as regular mail with probability 0.01. Bob checks his Inbox after a long time and selects a message at random. What is the probability it is spam?

Spam enters the Inbox folder as PP with rate 0.2, while regular email enters the Inbox folder as PP with rate 9.5. Viewing this is a combined PP of rate 9.7, the probability that a selected email is Spam is $2/97$.

(e) **[5]** Let (X, Y) be picked uniformly in the unit circle centered at $(0,0)$. What is $Q[Y|X]$ (the quadratic least squares estimate of Y given X)? *Hint.* Recall that the quadratic least squares estimate of Y given X is the quadratic function of X which minimizes the expected squared distance to Y among all quadratic functions of X: $E[(Y - Q[Y|X])^2] \leq E[(Y - aX^2 - bX - c)^2]$ for all $a, b, c \in \mathbb{R}$.

One has $Q[Y|X] = 0$ because $E[Y|X] = 0$, by symmetry.

(f) **[5]** Let X, Z be i.i.d. $\mathcal{N}(0,1)$ and $Y = X + Z$. Find $E[X^2|Y]$.

$E[X^2|Y] = E[X|Y]^2 + var(X|Y)$. $E[X|Y] = L[X|Y] = 0.5Y$ since X and Y are jointly Gaussian. Let $V = X - 0.5Y = 0.5(X - Z)$. Observe that V and Y are uncorrelated, and therefore independent since they are jointly Gaussian: $E[VY] = 0.5E[(X-Z)(X+Z)] = 0.5E[X^2 - Z^2] = 0$. $var(X|Y) = var(V + 0.5Y|Y) = var(V) = var(0.5(X - Z)) = 0.25(1+1) = 0.5$. Thus $E[X^2|Y] = 0.25Y^2 + 0.5$.

(g) **[5]** Let $X \in \{0,1\}$ and $Y = (X+1)Z$ where $Z = Expo(1)$ is independent of X. Find $MLE[X|Y = y]$.

One has $f_{Y|X}[y|0] = e^{-y}1\{y \geq 0\}$ and $f_{Y|X}[y|1] = 0.5e^{-0.5y}1\{y \geq 0\}$. Hence, $f_{Y|X}[y|0] = e^{-y}$ and $f_{Y|X}[y|1] = 0.5e^{-0.5y}$. Thus, $f_{Y|X}[y|1] > f_{Y|X}[y|0]$ if $0.5e^{-0.5y} > e^{-y}$, i.e., if $e^{0.5y} > 2$, or $0.5y > \ln(2)$. Hence, $MLE[X|Y] = 1\{Y > 2\ln(2)\}$.

(h) **[5]** Let X, Y be i.i.d. $\mathcal{N}(0,1)$. Find $Q[(X+2Y)^2|Y]$.

Since $Q[X|Y]$ has the form $aY^2 + bY + c$, it obeys linearity just as expectation does. One has $Q[X^2 + 2XY + 4Y^2|Y] = Q(X^2|Y) + 2YQ(X|Y) + 4Y^2$. Since X and Y are independent, this simplifies to $E(X^2) + 2YE(X) + 4Y^2 = 1 + 4Y^2$.

Problem 2 [30]

Part 1 [10].

The CTMC $\{X_t, t \geq 0\}$ has the transition diagram shown in Figure 2. Let $T_2 = \min\{t \geq 0 \mid X_t = 2\}$. Find $E[T_2|X_0 = 0]$.

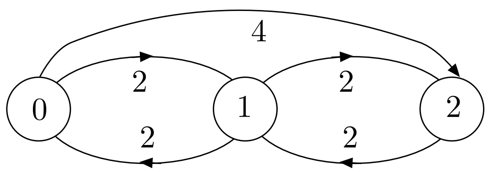

Figure 2: State transition diagram for Problem 2.

The Markov chain spends an exponentially distributed time with rate 6 in state 0, then jumps to 1 with probability $2/6$ or to 2 with probability $4/6$. When it is in state 1, it spends there an exponentially distributed time with rate 4 and then jumps to 0 with probability $1/2$ or to 2 with probability $1/2$. Thus, if we define $\beta(0) = E[T_2|X_0 = 0]$ and $\beta(1) = E[T_2|X_0 = 1]$, we see that

$$\begin{aligned} \beta(0) &= \frac{1}{6} + \frac{2}{6}\beta(1) + \frac{4}{6} \times 0 \\ \beta(1) &= \frac{1}{4} + \frac{1}{2}\beta(0) + \frac{1}{2} \times 0. \end{aligned}$$

Combining these equations, we find

$$\beta(0) = \frac{1}{6} + \frac{2}{6}[\frac{1}{4} + \frac{1}{2}\beta(0)] = \frac{1}{4} + \frac{1}{6}\beta(0).$$

Hence, $(5/6)\beta(0) = 1/4,$ so that $\beta(0) = 3/10$.

Part 2 [20].

Consider the three-state CTMC in Figure 3. The number on the edge directed from state i to state j is $q_{i,j}$, i.e., the transition rate from i to j. Assume that the process is in steady state, i.e., has its invariant distribution.

(a) [5] Find the long term time average fraction of time spent in state i for each i.

This is precisely $\pi(i) = \frac{1}{3}$ for all i.

(b) [5] Given that the process is in state i at time t, find the average time after time t until the process leaves state i, for each i.

The time the process spends on state i is exponential with rate $\lambda(i) = \sum_j q_{i,j}$, and hence the

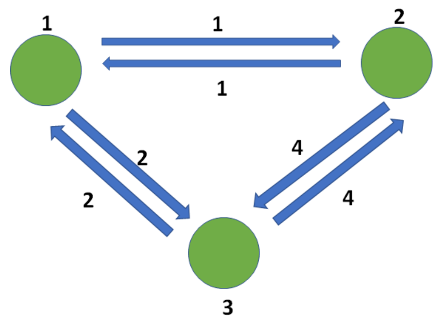

Figure 3: Rate transition diagram for Problem 2-B.

expected delay is $\frac{1}{\lambda(i)}$. For node 1, this is $\frac{1}{1+2}$, node 2, $\frac{1}{1+4}$, and node 3, $\frac{1}{2+4}$.

(c) [**5**] Find the long term time average fraction of transitions that go into state i, for each i.

Now consider the embedded Markov chain. From rate matrix, one can compute the transition probabilities of this chain. We try to find the stationary distribution of the chain, denoted by $\tilde{\pi}(i)$ for all i. Upon solving, we obtain $\tilde{\pi}(1) = \frac{3}{14}, \tilde{\pi}(2) = \frac{5}{14}, \tilde{\pi}(3) = \frac{3}{7}$.

(d) [**5**] Find the steady state probability that the next state to be entered is state 1.

In the steady state, the process is in state i with probability $\pi(i)$. Once in state 1, the next state is either 2 or 3 and cannot be 1. Once in state 2, the next state is 1 with probability $\frac{1}{5}$, similarly when in state 3, the next state is 1 with probability $\frac{1}{3}$. Hence the steady state probability that the next state is 1 is,

$$\pi(2) \times \frac{1}{5} + \pi(3) \times \frac{1}{3} = \frac{8}{45}.$$

Problem 3 [20]

Part 1 [10].

Let $X = B(p)$. and assume that $X' \in \{0,1\}$ is such that $P[X' = x'|X = x] = P(x, x')$ for $x, x' \in \{0,1\}$. Assume also that Y is such that $P[Y = y|X = x, X' = x'] = Q(x', y)$ for $x, x' \in \{0,1\}$ and $y \in \{1, \ldots, M\}$. Compute $P[X' = 1|Y = y]$. (*Hint:* Bayes' rule.)

$$P[X' = 1|Y = y] = \frac{P[X' = 1, Y = y]}{P(Y = y)}.$$

Now,

$$P[X' = 1, Y = y] = P(X' = 1)P[Y = y|X' = 1] = [(1-p)P(0,1) + pP(1,1)]Q(1,y).$$

Similarly,

$$P[X' = 0, Y = y] = P(X' = 0)P[Y = y|X' = 0] = [(1-p)P(0,0) + pP(1,0)]Q(0,y).$$

Consequently,

$$\begin{aligned} P(Y = y) &= P[X' = 1, Y = y] + P[X' = 0, Y = y] \\ &= [(1-p)P(0,1) + pP(1,1)]Q(1,y) + [(1-p)P(0,0) + pP(1,0)]Q(0,y). \end{aligned}$$

Finally,

$$P[X' = 1|Y = y] = \frac{[(1-p)P(0,1) + pP(1,1)]Q(1,y)}{[(1-p)P(0,1) + pP(1,1)]Q(1,y) + [(1-p)P(0,0) + pP(1,0)]Q(0,y)}.$$

Part 2 [**10**].

Consider a $HMC(\pi, P, Q)$ where P is on $\{0,1\}$. That is, X_n is a Markov chain on $\{0,1\}$ with transition matrix P and $P[Y_n = y|X_n = x] = Q(x,y)$. Let $\hat{X}_n = E[X_n|Y^n]$.

(a) [**4**] Explain clearly why one should be able to compute $\hat{X}_{n+1}$ as a function of $\hat{X}_n$ and Y_{n+1}. (*Hint:* Try to relate this problem to Part 1.)

This is the same problem as in Part 1 where given Y^n we know that $X_n = B(\hat{X}_n)$.

(b) [**6**] Derive the equations $\hat{X}_{n+1} = g(\hat{X}_n, Y_{n+1})$.

They are the same as in Part 1 where we replace p by $\hat{X}_n$. That is,

$$\hat{X}_{n+1} = \frac{[(1-\hat{X}_n)P(0,1) + \hat{X}_nP(1,1)]Q(1,y)}{[(1-\hat{X}_n)P(0,1) + \hat{X}_nP(1,1)]Q(1,y) + [(1-\hat{X}_n)P(0,0) + \hat{X}_nP(1,0)]Q(0,y)}.$$

Note that this recursive filter is not linear.

Problem 4 [**20**]

Consider the following dynamics equations (all random variables are zero-mean scalars):

$$\begin{aligned} &X_0 \sim \mathcal{N}(0, \sigma_0^2) \\ &X_1 = bX_0 + U, \\ &X_2 = a_0X_0 + a_1X_1 + W, \end{aligned}$$

where $U \sim \mathcal{N}(0, \sigma_U^2)$, $W \sim \mathcal{N}(0, \sigma_W^2)$, and (X_0, U, W) are independent.

(a) [**4**] What is the MMSE estimate of X_2 given X_0?

One has $L[X_2 \mid X_0] = a_0 X_0 + a_1 L[X_1 \mid X_0] = (a_0 + a_1 b) X_0$.

(b) **[4]** What is the MMSE estimate of X_2 given (X_0, X_1)?

Note that $L[X_2 \mid X_0, X_1] = a_0 X_0 + a_1 X_1$ since W is independent of (X_0, X_1).

(c) **[12]** Now, suppose that X_1 is replaced by an MMSE estimate $Y_1 := X_1 + V$, where $V \sim \mathcal{N}(0, \sigma_V^2)$ is independent of (X_0, U, W). What is the MMSE estimate of X_2 given (X_0, Y_1)?

From before, $L[X_2 \mid X_0] = (a_0 + a_1 b) X_0$. The innovation is

$$\tilde{Y}_1 := Y_1 - L[Y_1 \mid X_0] = Y_1 - L[X_1 + V \mid X_0] = Y_1 - bX_0.$$

So,

$$\begin{aligned} L[X_2 \mid \tilde{Y}_1] &= \frac{\operatorname{cov}(X_2, \tilde{Y}_1)}{\operatorname{var} \tilde{Y}_1} \tilde{Y}_1 = \frac{\operatorname{cov}((a_0 + a_1 b) X_0 + a_1 U + W, U + V)}{\operatorname{var}(U + V)} (Y_1 - bX_0) \\ &= \frac{a_1 \sigma_U^2}{\sigma_U^2 + \sigma_V^2} (Y_1 - bX_0). \end{aligned}$$

Thus,

$$L[X_2 \mid X_0, Y_1] = \Big(a_0 + a_1 b - \frac{a_1 b \sigma_U^2}{\sigma_U^2 + \sigma_V^2}\Big) X_0 + \frac{a_1 \sigma_U^2}{\sigma_U^2 + \sigma_V^2} Y_1 = \Big(a_0 + \frac{a_1 b \sigma_V^2}{\sigma_U^2 + \sigma_V^2}\Big) X_0 + \frac{a_1 \sigma_U^2}{\sigma_U^2 + \sigma_V^2} Y_1.$$

Comparing the answers to (b) and (c), one sees that certainty equivalence does not hold, that is, the optimal estimate changes when the observation X_1 is replaced by a noisy estimate Y_1. To see why, we can rewrite the dynamics equations in terms of Y_1:

$$\begin{aligned} &X_0 \sim \mathcal{N}(0, \sigma_0^2), \\ &Y_1 = bX_0 + \tilde{U} && \tilde{U} := U + V, \\ &X_2 = a_0 X_0 + a_1 Y_1 + \tilde{W} && \tilde{W} := W - a_1 V. \end{aligned}$$

The problem is that the noise variables $\tilde{U}$ and $\tilde{W}$ are no longer independent. In other words, replacing X_1 by the MMSE estimate Y_1 correlates the noise associated with $X_0 \to X_1$ and the noise associated with $(X_0, X_1) \to X_2$, which changes our optimal estimate.

Problem 5 [20]

Consider two independent random variables $X_1 \sim \mathcal{N}(\mu_1, 1)$, $X_2 \sim \mathcal{N}(\mu_2, 1)$ (where μ_1, μ_2 are unknown), and we would like to detect if $\mu_1 \neq \mu_2$ subject to the constraint that the probability of false alarm is at most α, where $\alpha \in (0, 1)$.

(a) [6] You are allowed to observe a linear combination of the two random variables:

$$Y := aX_1 + bX_2.$$

Explain clearly how you should choose a and b.

Since we are trying to detect whether the two means are different, then a good estimator of the difference of means is $X_1 - X_2$ (or any scalar multiple of $X_1 - X_2$).

Also observe that in order to control the probability of false alarm, we must know the distribution of Y under the null hypothesis ($\mu_1 = \mu_2$). However, under the null hypothesis, $aX_1 + bX_2 \sim \mathcal{N}(a\mu + b\mu, a^2 + b^2)$ (where μ denotes the common mean of X_1 and X_2), and since μ is unknown, the distribution of $aX_1 + bX_2$ is known only if $a = -b$.

(b) [14] Now suppose the null hypothesis is $\mu_1 = \mu_2$, and the alternate hypothesis is $\mu_1 = \mu_2 + \delta$, where $\delta > 0$ is known. You observe Y as before. Give the Neyman-Pearson decision rule to maximize the probability of correct detection with the constraint on the probability of false alarm.

Under the null hypothesis, $Y = X_1 - X_2 \sim \mathcal{N}(0, 2)$, and under the alternate hypothesis, $Y \sim \mathcal{N}(\delta, 2)$. The likelihood ratio is

$$L(y) := \frac{f(y \mid 1)}{f(y \mid 0)} = \frac{\exp(-(y-\delta)^2/4)}{\exp(-y^2/4)} = \exp\Big(\frac{\delta y}{2} - \frac{\delta^2}{4}\Big).$$

The likelihood ratio is increasing so the optimal test rejects for large values of Y, i.e., declare that $\mu_1 = \mu_2 + \delta$ if $Y > c$ for some cutoff c. We solve for c to satisfy the constraint on the probability of false alarm (there is no need for randomization here). Now, under the null hypothesis, $\mathbb{P}(Y > c) = \Phi(-c/\sqrt{2}) = \alpha$, so take $c := -\sqrt{2}\Phi^{-1}(\alpha)$.

Problem 6 [5 Bonus]

Thank you for taking the course! We hope you learned a lot and had fun along the way. Please let us know how the course went for you. What did you like and what did you dislike? Do you have any feedback for us?

We greatly appreciate your input.