CS 70 Discrete Mathematics and Probability Theory
Fall 2016 Seshia and Walrand Final Solutions

1. TRUE or FALSE?: 2x8=16 points

Clearly put your answers in the answer box on the bottom of this page!

This is what is to be graded. No need to justify!

Answer: Note that the answers provide explanations for your understanding, even though no such justification was required

1. A finite irreducible Markov chain that can go from state 1 to itself in one step is such that $P[X_n = 1]$ must converge as $n \to \infty$.

   Answer: True, because it is aperiodic.

2. We are given a probability space with sample space Ω and probability $P(\cdot)$. Let $\{A_1, A_2, \ldots, A_5\}$ be pairwise disjoint events whose union is Ω and let B be another event. If $P[B|A_1] = 1$, then $P[A_1|B] \geq P[A_k|B]$ for $k = 1, \ldots, 5$.

   Answer: False. For instance, it could be that $P(A_1) \ll 1$.

3. You throw 15 balls independently and uniformly at random into 7 bins. Then, the number of balls in bins 3 and 5 are positively correlated.

   Answer: False. They are negatively correlated.

4. The number of ways of distributing 10 identical smartphones to Steve, Bill, and Larry is $\binom{10}{3}$.

   Answer: False. Stars and bars, it should be $\binom{12}{2}$.

5. You draw five cards from a perfectly shuffled standard 52-card deck. The probability that the third card is an ace is strictly smaller than the probability that the first card is an ace.

   Answer: False. These probabilities are equal, by symmetry.

6. If $Q[Y|X] = 3 + 2X$, then $L[Y|X] = 3 + 2X$.

   Answer: True. The best linear estimate of Y cannot be better than the best quadratic estimate.

7. Assume that $E[(Y - 3\cos(5X) - 13\cos(7X))\cos(nX))] = 0$ for $n = 0, 1, \ldots, 9$. Then the values of $a_0, a_1, \ldots, a_9$ that minimize $E((Y - a_0 - a_1\cos(X) - a_2\cos(2X) - \cdots - a_9\cos(9X))^2)$ are such that $a_k = 0$ for $k \notin \{5, 7\}$ and $a_5 = 3, a_7 = 13$.

   Answer: True. This is the projection property.

8. Let p be a prime number. The set of all functions from $\{0, 1, 2, \ldots, p-1\}$ to $\mathbb{Q}$ is uncountably infinite.

   Answer: False. It is countably infinite. You can think of this set as isomorphic to the p-wise cartesian product $\mathbb{Q} \times \mathbb{Q} \times \ldots \times \mathbb{Q}$, and one can enumerate this set in a manner analogous to $\mathbb{Q}$ as covered in class.

Answer Box

(*Answer TRUE or FALSE in the corresponding box below.*)

1	2	3	4	5	6	7	8

2. Short Answers: 4x5=20 points
Provide a clear and concise justification of your answer.

1. You roll a six-sided balanced die until you get the first 6. How many dots do you accumulate, on average, including the 6 on the last roll?

   Answer: The average number of dots on a roll that does not yield a 6 is $(1+2+3+4+5)/5 = 3$. On average, you have to roll the die 6 times to get the first 6, thus five times before the first 6. You then accumulate, on average, $5 \times 3 + 6 = 21$ dots. (You may have noticed the use of Wald's identity.)

2. Alice, Bob, and Charles want to flip a coin to decide who pays for lunch, and want each to have the same probability of paying. However, they only have a biased coin and they don't even know its bias. Explain a mechanism that achieves the objective.

   Answer: You flip the coin three times. If the outcome has a single heads, then Alice pays if it is HTT, Bob pays if it is THT, and Charles pays if it is TTH. Otherwise, you ignore the outcome and you repeat the experiment.

3. You have to pass two courses in sequence, A then B, to be able to declare CS as your major. When you take a course for the first time, you pass it with probability $1/2$. If you take it for the second time, you pass it with probability $2/3$. If you fail a course twice, you cannot declare CS as your major. What is the probability that you can declare CS as your major?

   Answer: The probability that you fail course A twice in a row is $(1/2)(1/3) = 1/6$. Thus, the probability that you pass course A is $5/6$. Similarly, the probability that you pass course B is $5/6$. Thus, the probability that you can declare CS as your major is $(5/6)^2 = 25/36 \approx 70\%$.

4. You pick n points independently and uniformly at random inside a unit circle. Let Z be the distance of the furthest point to the center. What is $E(Z)$?

   Answer: We see that $P(Z < z) = P(Z_1 < z)^n$ where Z_1 is the distance from the first point to the center of the circle. But $P(Z_1 < z) = z^2$. Thus, $P(Z < z) = z^{2n}$, so that the pdf of Z is $2nz^{2n-1}1\{0 < z < 1\}$. Hence, $E(Z) = \int_0^1 z(2nz^{2n-1})dz = \frac{2n}{2n+1}$.

5. Suppose we draw cards from a standard 52-card deck without replacement. What is the expected number of cards we must draw before we get all 13 spades (including the last spade)?

   Answer: Define X to be the number of cards we draw before we get all 13 spades, including the last spade. Note that $X = 52 - Y$, where Y is the number of cards that come after all the spades. Arbitrarily label all the non-spade cards from 1 to 39. Define Y_i as the indicator variable that denotes whether the non-spade card with label i comes after all the spades. Then $Y = \sum_{i=1}^{39} Y_i$. Note that $E(Y_i) = \frac{1}{14}$. This is because if we consider only the relative orders of the 13 spades and the card labelled i, there is a $\frac{1}{14}$ chance that the card labelled i comes last. Applying linearity of expectation, we get

   $$E(X) = E(52 - Y) = 52 - E(Y) = 52 - 39\frac{1}{14} = \frac{689}{14}.$$

3. Rolling Dice I: 3+3+3+3=12 points

Provide a clear and concise justification of your answer.

In this problem, you roll two balanced six-sided dice.

1. What is the probability that the number of dots on the first die is at least twice the number on the second die?

   Answer: If the second die yields 1, then there are 5 possibilities for the first die, if it yields 2, there are 3, and if it yields 3, there is one. Thus, the probability is $(1/36)[5+3+1] = 1/4$.

2. What is the probability that the first die yields 4 if the total number of dots on the two dice is 8?

   Answer: It is $P[(4,4)|(2,6),(3,5),(4,4),(5,3),(6,2)] = 1/5$.

3. What is the probability that the sum is 8 given that the second die yields 4?
   Answer: It is the probability that the first die yields 4, i.e., $1/6$.
4. What is the probability that the number of dots on the first die is at least as large as on the second?
   Answer: Let X, Y be the number of dots on the two dice. It's easy to see that $P(X = Y) = 1/6$ so by symmetry $P(X > Y) = (1 - 1/6)/2 = 5/12$. Thus, $P(X \geq Y) = P(X > Y) + P(X = Y) = 5/12 + 1/6 = 7/12$.

4. Tossing Coins: 6+6= 12 points
Provide a clear and concise justification of your answer.

1. You are given two coins A and B. One is biased with $P(H) = 0.6$ and the second is fair. You don't know which is biased; it is equally likely to be A or B. You flip the pair of coins repeatedly (both each time) until at least one of the two coins yields H. It turns out that you have to flip them 27 times and that coin A then yields H while coin B yields T. What is the probability that A is the biased coin?
   Answer: Call F the event that we described. Then,
   $$P[F|A \text{ is biased}] = 0.4^{26}0.6 \times 0.5^{27} \text{ and } P[F|A \text{ is fair}] = 0.5^{27} \times 0.4^{27}.$$
   Thus, by Bayes' Rule,
   $$\begin{aligned} P[A \text{ is biased}|F] &= \frac{(1/2)P[F|A \text{ is biased}]}{(1/2)P[F|A \text{ is biased}] + (1/2)P[F|A \text{ is fair}]} \\ &= \frac{0.4^{26}0.6 \times 0.5^{27}}{0.4^{26}0.6 \times 0.5^{27} + 0.5^{27}0.4^{27}} = \frac{0.6}{0.6+0.4} = 0.6. \end{aligned}$$
2. You flip a fair coin 10 times. Let X be the total number of heads in the first 5 flips and Y the total number of heads in the odd flips, i.e., in flips $1, 3, 5, 7$, and 9. (a) Find $E[X|Y]$. (b) Find $E[Y|X]$.
   Answer: Let $Z_n = 1$ if coin n yields H. Then, $X = Z_1 + Z_2 + Z_3 + Z_4 + Z_5$ and $Y = Z_1 + Z_3 + Z_5 + Z_7 + Z_9$.
   (a) Now, $E[Z_1|Y] = Y/5$ by symmetry and $E[Z_2|Y] = 1/2$ by independence. Thus,
   $$E[X|Y] = (3/5)Y + 2(1/2) = 1 + (3/5)Y.$$
   (b) Similarly, $E[Z_1|X] = (1/5)X$ and $E[Z_9|X] = 1/2$. Thus,
   $$E[Y|X] = (3/5)X + 2(1/2) = 1 + (3/5)X.$$

5. Rolling Dice II: 8+6=14 points
Provide a clear and concise justification of your answer.

1. You roll a six-sided balanced die until either you get a 6 or the number of dots on a roll exceeds the number of dots on the previous roll. How many rolls do you need, on average? Write the equations that you have to solve to find the answer. Do not solve the equations.
   Answer: Let $\beta(n)$ be the average number of rolls you still need, given that the last roll yielded n dots and let β be the total average number of rolls. Thus $\beta(6) = 0$ and the first step equations are
   $$\begin{aligned} \beta &= 1 + (1/6)\sum_{n=1}^{5} \beta(n) \\ \beta(n) &= 1 + (1/6)\sum_{m=1}^{n} \beta(n), n = 1, \ldots, 5. \end{aligned}$$

2. You have a die and want to check if it is balanced. To do this, you decide to test the mean. You roll the die 10^4 times and find an average number of dots per roll equal to 3.2.

   (a) Using Chebyshev's inequality and a simple upper bound on the variance, find a 95% confidence interval for the mean. (*Hint: It can be shown that variance is bounded by 36.*)

   (b) Are you 95% sure that the die is not balanced?

   Answer:

   (a) We know that this interval is $3.2 \pm 4.5 \frac{\sigma}{\sqrt{n}} = 3.2 \pm 4.5 \times 10^{-2}\sigma$. It remains to find an upper bound on σ. Since $|X - E(X)| \leq 6$, we conclude that $var(X) \leq 36$, i.e., $\sigma \leq 6$. Hence, a 95% confidence interval is $3.2 \pm 4.5 \times 10^{-2} \times 6$, i.e., 3.2 ± 0.27.

   (b) Yes, because 3.5 is not in this interval.

6. Graphs and Chains: 8+6=14 points
Provide a clear and concise justification of your answer.

1. Imagine you are traversing a 3-dimensional hypercube. Beginning at the vertex 000, you repeatedly randomly pick any outgoing edge from your current vertex and follow it to another vertex. What is the expected number of edges you must traverse before reaching the vertex 111 for the first time?

   Answer: Our journey can be modeled as a Markov Chain with nodes 0, 1, 2, and 3, where node i is the collection of nodes with i ones in it. That is, node 0 consists of vertex 000; node 1 consists of vertices 001, 010, and 100; node 2 consists of vertices 110, 101, and 011; and node 3 consists of vertex 111. We want to find $E(X_0)$, where X_i denotes the number of edges we must traverse from node i to node 3. We can construct the following equations:

$$\begin{aligned} E(X_0) &= 1 + E(X_1) \\ E(X_1) &= 1 + \frac{2}{3}E(X_2) + \frac{1}{3}E(X_0) \\ E(X_2) &= 1 + \frac{2}{3}E(X_1) + \frac{1}{3}E(X_3) \\ E(X_3) &= 0 \end{aligned}$$

   Solving this linear system of equations, we get $E(X_0) = 10$.

2. Consider a Markov chain on $\{0,1\}$ with $P(0,1) = \alpha = P(1,0)$ for some $\alpha \in [0,1]$. Let $X_0, X_1, X_2, \ldots, X_n$ be the successive values of the Markov chain.

   (i) Propose a method to estimate α.

   (ii) Using Chebyshev's inequality, what is a 95% confidence interval for α.

   Answer:

   (i) We note that the Markov chain flips with probability α at each step, independently. Thus, we let $Y_m = 1\{X_{m-1} \neq X_m\} = |X_{m-1} - X_m|$ and we see that these are i.i.d. and equal to one with probability α and to zero otherwise. Hence, our estimate is

$$A_n = \frac{Y_1 + \cdots + Y_n}{n}.$$

   (ii) As we know by looking at coin flips, a 95% confidence interval for α is $A_n \pm \frac{2.25}{\sqrt{n}}$, using Chebyshev. (We can replace 2.25 by 1 for large n by the CLT, but you are not expected to remember this.)

7. Erdos-Renyi Random Graph: 4+2+4+2=12 points
Provide a clear and concise justification of your answer.

Random graph model is a key idea in network analysis. Erdos-Renyi Graph Model is the most popular and simplest network model. Suppose you have n vertices and you leave them unconnected at the beginning. Then for every two vertices, you draw an undirected edge with probability p. (p is a fixed number for all edges). Suppose we label all the vertices in an order 1, 2, ..., n.

1. Prove that in any E-R graph, if u is a vertex of odd degree in, then there exists a path from u to another vertex v of the graph where v also has odd degree. A *path* is a sequence of edges (v_1, v_2) ,(v_2, v_3), ..., (v_{n_2}, v_{n_1}), (v_{n_1}, v_n), where no vertex is repeated.

   Answer: We will build a path that does not reuse any edges. As we build the path, imagine erasing the edge we used to leave so that we will not use it again. Begin at vertex u and select an arbitrary path away from it. This will be the first component of the path. If, at any point, the path reaches a vertex of odd degree, we will be done, but each time we arrive at a vertex of even degree, we are guaranteed that there is another vertex out, and, having left, we effectively erase two edges from that meet at the vertex. Since the vertex originally was of even degree, coming in and going out reduces its degree by two, so it remains even. In this way, there is *always* a way to continue when we arrive at a vertex of even degree. Since there are only a finite number of edges, the tour must end eventually, and the only way it can end is if we arrive at a vertex of odd degree.

2. What's the probability that there are 0 paths of length 2 connecting vertices 1 and 2?

   Answer: Let $I(v, w)$ be the indicator that there is an edge connecting the vertices v and w. Using independence of the edges we compute:

$$\prod_{v=3}^{n}(1 - P[(I(1, v) = 1)I(v, 2) = 1)]) = (1 - p^2)^{(n-2)}$$

3. Let X_l be the number of paths of length l connecting vertices 1 and 2. What's the expectation and variance of X_2?

   Answer: We say that there exists a path of length l connecting two given vertices $v \neq w$ if there exist vertices $v_1, \dots, v_{l-1}$ such that $v \sim v_1 \sim v_2 \sim \cdots \sim v_{l-1} \sim w$, where $v \sim w$ means that there is an edge between v and w.

   By definitions we have

$$X_2 = \sum_{v=3}^{n} \mathbf{1}(1 \sim v, v \sim 2)$$

   Taking expectation yields

$$E(X_2) = \sum_{v=3}^{n} P(1 \sim v)P(v \sim 2) = (n-2)p^2$$

   Now we compute $E(X_2^2)$:

$$\begin{aligned} E(X_2^2) &= \sum_{v,w=3}^{n} E[\mathbf{1}(1 \sim v \sim 2)\mathbf{1}(1 \sim v \sim 2)] \\ &= \sum_{v=3}^{n} P(1 \sim v)P(v \sim 2) + \sum_{v,w=3:v\neq w}^{n} P(1 \sim v)P(v \sim 2)P(1 \sim w)P(w \sim 2) \\ &= (n-2)p^2 + (n-2)(n-3)p^4 \end{aligned}$$

So the variance of X_2 becomes:

$$\begin{aligned}\text{Var}(X_2) &= E(X_2^2) - E(X_2)^2 \\ &= (n-2)p^2 + (n-2)(n-3)p^4 - [(n-2)p^2]^2 \\ &= (n-2)(1-p^2)p^2\end{aligned}$$

4. What's the expectation of X_l? (You should give an expression in terms of n, p and l and you could leave a summation, integral or product if that's easier to write the final answer.)

Answer: By definitions we have

$$X_l = \sum \mathbf{1}(1 \sim v_1, v_1 \sim v_2, \dots, v_{l-1} \sim 2)$$

where the sum is over ordered but distinct vertices v_i taken from the set $\{3, \dots, n\}$. Let's first think about how many ways to choose $l+1$ vertices (to form a path of length l) from n: $\binom{n}{l+1} = n(n-1)\cdots(n-l)$. The situation in this problem is slightly different where you need to fix the first two vertices to be 1 and 2 respectively. So the total number of ways to choose vertices from n becomes $(n-2)(n-3)\cdots(n-l)$.

Taking expectation yields

$$\begin{aligned}E(X_l) &= \sum P(1 \sim v_1, v_1 \sim v_2, \dots, v_{l-1} \sim 2) \\ &= \sum P(1 \sim v_1)P(v_1 \sim v_2)\cdots P(v_{l-1} \sim 2) \\ &= (n-2)(n-3)\dots(n-l)p^l\end{aligned}$$

8. Polling for 2020: 10 points Provide a clear and concise justification of your answer.

We poll 100 men and 200 women. Among those, 30 men and 160 women declare they would vote for Michelle Obama for President in 2020. Assume that, among the voters in the general election, there will be an equal number of men and women and that your poll is representative of the voting in the general election. What is your 95% confidence interval for the fraction of people who will vote for Michelle? Would you advise her to run? Justify your answer.

To solve this problem, we suggest that you follow the following steps: (1) What is your estimate of the fraction of the people who will vote for Michelle in the general election? (2) What is its variance?; (3) What is an upper bound on that variance?; (4) Write down Chebyshev and use the upper bound; (5) Choose the distance from the mean so that the Chebyshev bound is 5%; (6) What is the resulting confidence interval?

Answer:

Let X_n be the indicators that men vote for Michelle and Y_n for women. Thus, we know that

$$V := \sum_{n=1}^{100} X_n = 30 \text{ and } W := \sum_{n=1}^{200} Y_n = 160.$$

(1) The fraction of people voting for Michelle is $p := E(X_1 + Y_1)/2$. Thus, a good estimate is $Z := V/200 + W/400$ since $E(Z) = p$. Now,

(2) One has

$$\sigma^2 := var(Z) = \frac{1}{(200)^2} var(V) + \frac{1}{(400)^2} var(W) = \frac{1}{400} var(X_1) + \frac{1}{800} var(Y_1).$$

(3) Since $var(X_1) \leq 1/4$ and $var(Y_1) \leq 1/4$, one has

$$\sigma^2 \leq (\frac{1}{400} + \frac{1}{800})\frac{1}{4} = \frac{3}{3200}.$$

(4) According to Chebyshev,

$$P(|Z - E(Z)| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \leq \frac{3}{3200\varepsilon^2}.$$

(5) If we want this probability to be 5%, we need

$$\frac{3}{3200\varepsilon^2} = 5\% = \frac{1}{20},$$

so that $\varepsilon^2 = 20\frac{3}{3200} = \frac{3}{160} \leq \frac{1}{50} = 2\%$.

(6) Thus, a 95% confidence interval is $Z \pm 2\% = (0.30 + 0.80)/2 \pm 2\% = 0.55 \pm 0.02 = [0.53, 0.57]$.

Yes, I would advise Michelle to run. The probability she would win is larger than 95%.

9. Random Variables: 3+6+6+8+7=30 points
Provide a clear and concise justification of your answer.

1. Define 'random variable' in one sentence. (No need for a justification.)

   Answer: A random variable is a real-valued function of the outcome of a random experiment.

2. Let $X_n, n \geq 1$ be i.i.d. with mean μ and variance σ^2. Let also $A_n = (X_1 + \cdots + X_n)/n$. (a) Using Chebyshev, find a 90% confidence interval for μ. (b) Apply this result to the situation where the X_n are 1 when a coin yields H and zero otherwise.

   Answer:

   (a) We have

   $$P(|A_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

   We choose ε so that the right-hand side is 0.1. This gives

   $$\frac{\sigma^2}{n\varepsilon^2} = 0.1, \text{ i.e., } \varepsilon^2 = 10\frac{\sigma^2}{n}.$$

   Thus, $\varepsilon = \sqrt{10}\sigma/\sqrt{n}$. The 90% confidence interval is then $A_n \pm \sqrt{10}\frac{\sigma}{\sqrt{n}}$.

   (b) For coin flips, $\mu = p$ and $\sigma \leq 1/2$. Hence, the 90% confidence interval for p is $A_n \pm \frac{\sqrt{10}}{2\sqrt{n}} \approx A_n \pm \frac{1.6}{\sqrt{n}}$.

3. You choose a point (X, Y) uniformly at random in the triangle with vertices $(-1, 0), (0, 1), (1, 0)$. (a) Find $E[Y|X]$. (b) Find $L[Y|X]$. (c) Are X and Y positively-, negatively-, or un-correlated?

   Answer:

   (a) Given that $X = x$, we know that $Y = U[0, 1 - |x|]$. Hence, $E[Y|X] = (1/2)(1 - |X|)$.

   (b) By symmetry, $L[Y|X]$ is constant. Therefore, it must be equal to $E[Y]$. Now, $P(Y > y) = (1 - y)^2$, since this happens when (X, Y) falls in a triangle with base $2(1 - y)$ and height $(1 - y)$. Thus, $f_Y(y) = 2(1 - y)$, so that $E[Y] = \int_0^1 2y(1 - y)dy = 1/3$, as we might have guessed. Hence, $L[Y|X] = 1/3$.

   (c) Since LLSE is a constant, clearly X and Y are uncorrelated. Or we can do a computation. We know that $E((Y - L[Y|X])X) = 0$, so that $E(XY) = E(XL[Y|X]) = (1/3)E[X] = 0$. Thus, $E(XY) = 0 = E(X)E(Y)$ and X, Y are uncorrelated.

4. Let Y and Z be independent random variables where Y is Poisson with mean 60 and Z is Poisson with mean 10. (a) Find $L[Y|X]$ where $X = Y + Z$. (b) Find $E[Y|X]$. (*Hint:* Recall that if V is Poisson with mean λ, then $E(V) = \lambda$ and $var(V) = \lambda$. Also, recall that the sum of Poisson independent random variables is Poisson.)

   Answer:

   (a) We know that

$$L[Y|X] = E(Y) + \frac{cov(X,Y)}{var(X)}(X - E(X)).$$

   Now, $E(Y) = 60$ and $E(X) = E(Y) + E(Z) = 70$. Also, $var(X) = var(Y) + var(Z) = 70$. Finally,

$$cov(X,Y) = cov(Y+Z,Y) = var(Y) = 60.$$

   Hence,

$$L[Y|X] = 60 + \frac{60}{70}(X - 70) = \frac{6}{7}X.$$

   (b) Let $Z, Z_1, \ldots, Z_6$ be independent and Poisson with mean 10. We can write $Y = Z_1 + \cdots + Z_6$ because this sum is Poisson with mean 60 and independent of Z. Then, $X = Z_1 + \cdots + Z_6 + Z$ and one has

$$E[Y|X] = E[Z_1 + \cdots + Z_6 | Z_1 + \cdots + Z_6 + Z] = \frac{6}{7}(Z_1 + \cdots + Z_6 + Z) = \frac{6}{7}X,$$

   by symmetry.

5. Let X be a discrete random variable that takes values in $[0,1]$. Show that its variance is at most $1/4$. *Hint:* Show that the random variable that takes at most n different values in $(0,1)$ and that has the largest variance is Bernoulli $1/2$. To do this, argue by contradiction and assume that X takes a value x in $[0.5,1)$ with probability p. Replace that value by 1 with probability $1/2$ and $2x-1$ with probability $1/2$. Show that the variance increases.

   Answer: The mean does not change. It suffices to show that the mean value of the square increases. This is the case because in the calculation of $E(X^2)$ the term x^2p gets replaced by $(p/2)(1)^2 + (p/2)(2x-1)^2$. But $x^2 < 1/2 + (1/2)(2x-1)^2$.

(Scratch space)