CS 70 Discrete Mathematics and Probability Theory
Fall 2017 Rao

Final

PRINT Your Name: ________________________________, ________________________________
(Last) (First)

READ AND SIGN The Honor Code:
As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. __________

PRINT Your Student ID: ______________________________________

WRITE your exam room:

WRITE the name of the person sitting to your left: ______________________________________

WRITE the name of the person sitting to your right: ______________________________________

PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY

- After the exam starts, please *write your student ID on every page*. **You will not be allowed to write** *anything* **once the exam ends**.
- We will not grade anything outside of the space provided for a problem unless we are clearly told in the space provided for the question to look elsewhere.
- The questions vary in difficulty. If you get stuck on any one, it helps to leave it and try another one.
- In general, no justification on short answer/true false questions unless otherwise indicated.
- Calculators are not allowed. **You do NOT need to simplify any probability related answers to a decimal fraction,** but your answer must be in the simplest form (no summations or integrals).
- You may consult only *3 sheets of notes*. Apart from that, you are not allowed to look at books, notes, etc. Any electronic devices such as phones and computers are NOT permitted.
- Regrades will be due quickly so watch piazza.
- There are **16** single sided pages on the exam. Notify a proctor immediately if a page is missing.
- You have **180** minutes: there are **7** questions on this exam worth a total of **215** points.

Do not turn this page until your proctor tells you to do so.

1. Discrete Math: True/False (2pts/part,10 parts. 20 points)

1. (True/False) If $d|m$ and $d|n$ then $d|(m-n)$. (Recall $x|y$ means "the integer x divides the integer y".)

   ◯ True

   ◯ False

2. (True/False) If $d|mn$ then $d|n$ or $d|m$.

   ◯ True

   ◯ False

3. (True/False) $\neg P \implies \neg Q$ implies that $\neg Q \implies \neg P$.

   ◯ True

   ◯ False

4. (True/False) Recall a tournament graph on n vertices has an directed edge between every pair of vertices in exactly one direction. If the tournament has a directed cycle of length 5, there is a smaller directed cycle.

   ◯ True

   ◯ False

5. (True/False) Given a 3-colorable n vertex graph, there exists a way to add a vertex and edges to the graph where the new vertex has degree $\lceil 2n/3 \rceil$ and still have a 3-colorable graph.

   ◯ True

   ◯ False

6. (True/False) If there are two different stable pairings in stable marriage instance, then it cannot be the case that all men have the same preference list.

   ◯ True

   ◯ False

7. (True/False) In a broken run of the traditional marriage algorithm where exactly one woman accidentally rejects a man but ends up with a man she likes better. The resulting pairing is stable.

   ◯ True

   ◯ False

8. (True/False) If men ask women in reverse order of preference one still gets a stable pairing, but this time it is female optimal.

   ◯ True

   ◯ False

9. (True/False) The length of every cycle in the hypercube is even.

   ◯ True

   ◯ False

2. Discrete Math:Short Answer (3 pts/part, 16 parts. 48 points.

1. $\neg(\forall x, P(x) \vee Q(x)) \equiv \exists x,$_________. (Fill in the blank.) (Make sure the negation is fully distributed.)

2. What is the size of $\{a \pmod n, 2a \pmod n, 3a \pmod n, \ldots, (n-1)a \pmod n\}$ if $gcd(a,n) = 1$?

3. What is $2^{16} \pmod 7$?

4. What is the size of the set $\{ay \pmod{pq} : y \in \{1, \ldots, pq-1\}\}$ when a is not a multiple of p or q?

5. What is the size of the set $\{ay \pmod{pq} : y \in \{1, \ldots, pq-1\}\}$ when a is a multiple of p (but not q)?

6. What is $5^{60} \pmod{77}$?

7. What is $7^{60} \pmod{77}$?

8. What is the minimum number of degree 1 vertices in an n-vertex tree for $n > 1$? (Answer could be an expression that involves n.)

9. What is the maximum number of degree 1 vertices in an n-vertex tree? (Answer could be an expression that involves n.)

10. What is the maximum number of edges in any simple planar graph with 5 vertices?

11. Consider Professor Rao's public key (N,e) and secret key d, which has 512 bits. Professor Rao wants to share the secret key with his three children where any two can recover the secret d.

    (a) What degree polynomial should he use?

    (b) How big should the field over which we are working be? (That is, how big should the modulus be for the modular arithmetic that we use.)

12. What is an error polynomial for the Berlekamp-Welsh method where the corrupted packets correspond to points with x-values $0,1$ and 3 working modulo 11?

13. For degree (at most) d non-zero distinct polynomials, $P(x)$ and $Q(x)$, what is the maximum number of roots that $P(x)-Q(x)$ can have?

14. We have to assign 750 students to rooms in CS70.

    (a) How many ways are there to do this in 3 rooms of capacity 240, 250, and 260?

    (b) How many ways are there to do this in 3 rooms of capacity 250, 260, and 270? (Notice there will be a total of 30 empty seats in this case.) (An expression that may involve summations.)

3. A Quick Proof. (12 points.)

You have n coins $C_1, C_2, \ldots, C_n$ for $n \in \mathbb{N}$. Each coin is weighted differently so that the probability that coin C_i comes up heads is $\frac{1}{2i+1}$. Prove by induction that if the n coins are tossed, then the probability of getting an odd number of heads is $\frac{n}{2n+1}$.

1. Base case.

2. State your induction hypothesis.

3. Do the inductive step.

4. Probability:True/False. (2pts/parts,6 parts. 12 points)

1. (True/ False) If false give a counterexample in the space provided next to the true-false bubbles. The example is graded.

   (a) If X,Y are independent, then $cov(X,Y) = 0$.

   ◯ True

   ◯ False

   (b) If $cov(X,Y) = 0$, then X,Y are independent.

   ◯ True

   ◯ False

2. (True/False) If X_1 and X_2 are i.i.d. Exp(1) random variables, $cov(\min(X_1,X_2),\max(X_1,X_2)) = 0$.

   ◯ True

   ◯ False

3. (True or False) If $X \sim Geom(p)$, then $E[X+m \mid X > n] = m+n+E[X]$.

   ◯ True

   ◯ False

4. The CLT can be used to bound the probability that a random variable is far from its mean.

   ◯ True

   ◯ False

5. Quick conceptual questions. (4 pts part, 3 parts. 12 points.)

1. Explain in words what it means when the covariance between two random variables X and Y is (a) positive (b) negative (c) zero.

2. In a sentence or two describe: (i) the difference between Bayesian and non-Bayesian linear regression: (ii) how you can study both perspectives using a single framework.

3. A sequence of random variables $X_0, X_1, X_2, X_3, \ldots$ is a Markov chain if: (fill in the blank)

SID:

6. Probability: Short Answer. 3pts/part. 32 parts. 96 pts.

1. In a class of 24 students, what is the probability that at least two students have the same birthday? (Assume the number of days in a year is 365. Answer is an expression, possibly with products. No need to simplify.)

2. Two real numbers are chosen uniformly from the unit interval. What is the probability that their sum is less than or equal to 1 given that one of them is less than or equal to 1/2?

3. If X,Y are independent continuous-valued random variables uniform in $[0,1]$. What is $E[X \mid X+Y = 1.5]$?

4. If $X_1, X_2, \ldots, X_n$ are i.i.d. $U[0,1]$ RVs.
   (a) Find the pdf of $Y = \min\{X_1, X_2, \ldots, X_n\}$.

   (b) Let $Z = \max\{X_1, X_2, \ldots, X_{100}\}$. What is $E[Z]$?

5. I want to take a student poll to find the popularity of EECS 70 (assume each student independently likes it with probability p), and I need to pay each student \$1 to get his/her opinion. Suppose I want to estimate p within 1 percent accuracy with a 95% confidence level, I want to find how much money I need to find my estimate.

   (a) What estimator could you use for p from a set of samples, $X_1, X_2, \ldots, X_n$? (It should have expectation p.)

   (b) What is an upper bound on the variance of your estimator that does not depend on p?

   (c) How much money would I need to spend if I use the CLT?

   (d) How about if I use the Chebyshev bound?

6. A relatively rare disease afflicts 1 in 100 people in the population. Screening for the disease has a missed detection rate of 1% (i.e. there is a 1% chance that a person has the disease but the test doesn't catch it), and a false alarm rate of 5% (i.e. there is a 5% chance that the test comes out positive for the disease even though the person does not have it), then if a test comes out positive, what is the probability that the person has the disease? (The answer can be an expression with numbers, no need to simplify into a single number.)

7. Let X, Y be a pair of random variables. The value of c that minimizes the variance of $X - cY$ is what? (You may refer to any of $E[X], E[Y], Cov(X,Y), Var(X)$, or $Var(Y)$ in your solution.)

8. Let X, Y, Z be i.i.d. $U[-1,1]$ RVs.

   (a) What is $E[X]$?

   (b) Find $E[(X+Y+Z)^2 \mid X = x]$.

9. The local Safeway has an essentially limitless number of cereal boxes, with each cereal box containing a tiny Marvel Comic superhero figure in it. You win a prize if you can collect 20 distinct superhero figures. Assume that there are a total of 20 distinct superheroes in the collection, with each box equally likely to contain any superhero.

   (a) What is the expected number of cereal boxes you need to buy to win the prize?

   (b) Now suppose that you have only a \$20 budget on the cereal boxes, and each cereal box costs \$1. What is the expected number of superheroes you will get with your \$20?

   (c) What is the variance of the number of distinct superheros that you collect with your \$20?

10. X and Y are continuous random variables and are uniformly distributed with pdf $f(x,y) = c$ over their region of support. Their region of support is the following: $\{1 < X < 2, 1 < Y < 4\} \cup \{2 < X < 3, 2 < Y < 3\}$.

 (a) Find c.

 (b) Find the marginal distributions of X and Y.

 (c) Find the MMSE (minimum least squares error) estimate of Y given X. (This should not take long, if you don't see it, move on.)

11. There are N passengers boarding a full flight. They have assigned seats but they have all lost their boarding passes, so they choose to sit in random seats (I know, in real life, they will probably get the aisle or window seats at the front of the plane, but we wanted to keep things simple for you).

    (a) What is the expected number of passengers who sit in their assigned seats?

    (b) What is the probability that i passengers sit in their assigned seats?

12. The lifespans of good batteries are exponentially distributed with mean 2 days. Those of used batteries are exponentially distributed with mean 1 day. There two batches of batteries, one batch has all new batteries, and the other batch has all used batteries, but you don't know which is which. You randomly select one of them and test one battery from the batch, and find that it lasts for 0.75 day.

    (a) Let p be the probability that you picked the batch of new batteries. What is p?

    (b) What is the expected lifespan of another battery in the batch you picked? (Leave your answer in terms of p, the answer to part (a).)

13. Let $X \sim expo(\lambda)$, and let $\lceil X \rceil$ denote the ceiling of X - that is, the smallest integer greater than or equal to X. Find the distribution of $\lceil X \rceil$ and identify this distribution as one we have learned in class, with appropriate parameters.

14. Let $X = N(0,1)$ and $Y = N(1,1)$ be independent Gaussian random variables. You get an observation $z = 0.6$ that is equally likely to be a realization of either X or Y. You want to decide whether z came from X or Y by evaluating which decision leads to a larger probability of being right.

    (a) If you decide z came from X, what is the probabilty that you are right? (No need to evaluate numerically, leave it as a function of the pdf of a standard Normal random variable).

    (b) Should you decide z is from X or from Y to get a larger probability of being right?

15. A hard-working GSI is holding her office hours (OH) for EECS 70 students. A random number of students enter and leave her office during her OH. Let us break up time into 1-minute segments, and assume that there is either 0 or 1 student in each time segment, and that this discrete arrival/departure process is well modeled by a 2-state Markov Chain. For each time segment, the transition probabilities are 0.8 for going from 0 students to 1 student in the OH, and 0.4 for going from 1 student to 0 students in the OH.

    (a) If the GSI starts off her office hours at $t = 0$ with 1 student, what is the probability that she has 0 student at time $t = 2$?

    (b) What does the probability go to, as t gets large, that there is 1 student?

16. You flip a fair coin repeatedly until you get 2 Heads in a row or 2 Heads in 3 tosses. We wish to find the expected number of tosses you need before you stop.

    (a) Draw a four state Markov Chain corresponding to this process, where the goal state is terminal (i.e., only has transitions to itself.)

    (b) What is the expected number of tosses you need to stop? (No need to solve the problem numerically, just set up the equations needed to solve it.)

7. Probability: Basketball. 15 points.

Alice (A) and Bob (B) play a one-on-one pickup game of basketball. Each made basket counts as 1 point. A beats B by a score of 51-49. We want to find the probability that A leads from start to finish (i.e. the game is never tied at any time other than at the start of the game) but we will lead up to this with some helpful hint questions that should help you find the answer.

Let us first define the following useful events:

T_A : Game had at least one tie and A got the first point;

T_B : Game had at least one tie and B got the first point;

N: Game had no tie..

1. How are $P(T_A), P(T_B)$, and $P(N)$ related?

2. Given the final score what is the probability that B got the first point?

3. How is $P(T_B)$ related to the probability that B got the first point?

4. How are $P(T_A)$ and $P(T_B)$ related? (Hint: use symmetry)

5. Using the above, What is the probability that A leads from start to finish (i.e. the game is never tied except at the start of the game)?