Section 10

Review

You probably want to look over this review sheet, with the caveat that there are quite a few things here that we have not covered this quarter.

Task 1 – Short answer

a) Consider a set S containing k distinct integers. What is the smallest k for which S is guaranteed to have 3 numbers that are the same mod 5 (in other words, for every pair of elements a and b in the set S, $a \bmod 5 = b \bmod 5$)?

b) Let X be a discrete random variable that can only be between -10 and 10. That is, $P(X = x) \geqslant 0$ for $-10 \leqslant x \leqslant 10$, and $P(X = x) = 0$ otherwise. What is the smallest possible value the variance of X can take?

c) How many ways are there to rearrange the letters in the word KNICKKNACK?

d) I toss n balls into n bins uniformly at random. What is the expected number of bins with exactly k balls in them?

e) Consider a six-sided die where $Pr(1) = Pr(2) = Pr(3) = Pr(4) = 1/8$ and $Pr(5) = Pr(6) = 1/4$. Let X be the random variable which is the square root of the value showing. (For example, if the die shows a 1, X is 1, if the die shows a 2, X is $\sqrt{2}$, if the die shows a 3, $X = \sqrt{3}$ and so on.) What is the expected value of X? (Leave your answer in the form of a numerical sum; do not bother simplifying it.)

f) A bus route has interarrival times (the times between subsequent arrivals) at a bus stop that are exponentially distributed with parameter $\lambda = \frac{0.05}{\text{min}}$. What is the probability of waiting an hour or more for a bus?

g) How many different ways are there to select 3 dozen indistinguishable colored roses if red, yellow, pink, white, purple and orange roses are available?

h) Two identical 52-card decks are mixed together. How many distinct permutations of the 104 cards are there?

Task 2 – Random boolean formulas

Consider a boolean formula on n variables in 3-CNF, that is, conjunctive normal form with 3 literals per clause. This means that it is an "and" of "ors", where each "or" has 3 literals. Each parenthesized expression (i.e., each "or" of three literals) is called a clause. Here is an example of a boolean formula in 3-CNF, with $n = 6$ variables and $m = 4$ clauses.

$$(x_1 \vee x_3 \vee x_5) \wedge (\neg x_1 \vee \neg x_2 \vee x_6) \wedge (x_5 \vee \neg x_3 \vee x_4) \wedge (\neg x_1 \vee x_4 \vee x_5).$$

a) What is the probability that $(\neg x_1 \vee \neg x_2 \vee x_3)$ evaluates to true if variable x_i is set to true with probability p_i, independently for all i?

b) Consider a boolean formula in 3-CNF with n variables and m clauses, where the three literals in each clause refer to distinct variables. What is the expected number of satisfied clauses if each variable is set to true independently with probability 1/2? A clause is satisfied if it evaluates to true. (In the displayed example above, if $x_1, \ldots, x_5$ are set to true and x_6 is set to false, then all clauses but the second are satisfied.)

Task 3 – Biased coin flips

We flip a biased coin with probability p of getting heads until we either get heads or we flip the coin three times. Thus, the possible outcomes of this random experiment are $< H >, < T, H >, < T, T, H >$ and $< T, T, T >$.

a) What is the probability mass function of X, where X is the number of heads. (Notice that X is 1 for the first three outcomes, and 0 in the last outcome.)

b) What is the probability that the coin is flipped more than once?

c) Are the events "there is a second flip and it is heads" and "there is a third flip and it is heads" independent? Justify your answer.

d) Given that we flipped more than once and ended up with heads, what is the probability that we got heads on the second flip? (No need to simplify your answer.)

Task 4 – Bitcoin users

There is a population of n people. The number of Bitcoin users among these n people is i with probability p_i, where, of course, $\sum_{0 \leqslant i \leqslant n} p_i = 1$. We take a random sample of k people from the population (without replacement). Use Bayes Theorem to derive an expression for the probability that there are i Bitcoin users in the population conditioned on the fact that there are j Bitcoin users in the sample. Let B_i be the event that there are i Bitcoin users in the population and let S_j be the event that there are j Bitcoin users in the sample. Your answer should be written in terms of the p_ℓ's, i, j, n and k. Your answer can contain summation notation.

Task 5 – Investments

You are considering three investments. Investment A yields a return which is X dollars where X is Poisson with parameter 2. Investment B yields a return of Y dollars where Y is Geometric with parameter 1/2. Investment C yields a return of Z dollars which is Binomial with parameters $n = 20$ and $p = 0.1$. The returns of the three investments are independent.

a) Suppose you invest simultaneously in all three of these possible investments. What is the expected value and the variance of your total return?

b) Suppose instead that you choose uniformly at random from among the 3 investments (i.e., you choose each one with probability 1/3). Use the law of total probability to write an expression for the probability that the return is 10 dollars. Your final expression should contain numbers only. No need to simplify your answer.

Task 6 – Another continuous r.v.

The density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & \text{when } 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbb{E}[X] = \frac{3}{5}$, find a and b.

Task 7 – Point on a line

A point is chosen at random on a line segment of length L. Interpret this statement (i.e., define the relevant random variable(s)) and find the probability that the ratio of the shorter to the longer segment is less than $\frac{1}{4}$.

Task 8 – Min and max of i.i.d. random variables

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables each with CDF $F_X(x)$ and pdf $f_X(x)$. Let $Y = \min(X_1, \ldots, X_n)$ and let $Z = \max(X_1, \ldots, X_n)$. Show how to write the CDF and pdf of Y and Z in terms of the functions $F_X(\cdot)$ and $f_X(\cdot)$.

Task 9 – CLT example

Let X be the sum of 100 real numbers, and let Y be the same sum, but with each number rounded to the nearest integer before summing. If the roundoff errors (the difference between a real number and that number rounded to the nearest integer) are independent and uniformly distributed between -0.5 and 0.5, what is the approximate probability that $|X - Y| > 3$?

Task 10 – Tweets

A prolific twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).

Task 11 – Will I Get My Package

A delivery guy in some company is out delivering n packages to n customers, where $n \in \{2, 3, 4, ..., \infty\}, n > 1$. Not only does he hand each customer a package uniformly at random from the remaining packages, he opens the package before delivering it with probability $\frac{1}{2}$. Let X be the number of customers who receive their own packages unopened.

a) Compute the expectation $\mathbb{E}[X]$.

b) Compute the variance $\text{Var}(X)$.

Task 12 – Subset Card Game

Jonathan and Yiming are playing a card game. The cards have not yet been dealt from the deck to their hands. This deck has $k > 2$ cards, and each card has a real number written on it. In this deck, the sum of the card values is 0, and that the sum of squares of the values of the cards is 1. Specifically, if the card values are $c_1, c_2, \ldots, c_k$, then we have $\sum_{i=1}^{k} c_i = 0$ and $\sum_{i=1}^{k} c_i^2 = 1$.

The cards are then going to be dealt randomly in the following fashion: for each card in the deck, a fair coin is flipped. If the coin lands heads, then the card goes to Yiming, and if the coin lands tails, the card goes to Jonathan. Note that it is possible for either player to end up with no cards/all the cards.

Calculate $\mathbb{E}[S]$ and $\text{Var}(S)$, where S is the sum of value of cards in Yiming's hand (where an empty hand corresponds to a sum of 0). The answer should not include a summation.

Task 13 – Random Variables Warm-Up

[Credit: Berkeley CS 70] Let X and Y be random variables, each taking values in the set {0,1,2}, with joint distribution

$$\begin{array}{rrr} \mathbb{P}[X=0,Y=0]=1/3 & \mathbb{P}[X=0,Y=1]=0 & \mathbb{P}[X=0,Y=2]=1/3 \\ \mathbb{P}[X=1,Y=0]=0 & \mathbb{P}[X=1,Y=1]=1/9 & \mathbb{P}[X=1,Y=2]=0 \\ \mathbb{P}[X=2,Y=0]=1/9 & \mathbb{P}[X=2,Y=1]=1/9 & \mathbb{P}[X=2,Y=2]=0 \end{array}$$

a) What are the marginal distributions of X and Y?

b) What are $\mathbb{E}[X]$ and $\mathbb{E}[Y]$?

c) Let I be the indicator that $X = 1$, and J be the indicator that $Y = 1$. What are $\mathbb{E}[I]$, $\mathbb{E}[J]$ and $\mathbb{E}[IJ]$?

d) In general, let I_A and I_B be the indicators for events A and B in a probability space $(\Omega, \mathbb{P})$. What is $\mathbb{E}[I_A I_B]$, expressed in terms of the probability of some event?

Task 14 – Joint Distributions

a) Give an example of discrete random variables X and Y with the property that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$. Specify the joint distribution of X and Y.

b) Give an example of discrete random variables X and Y that (i) are *not independent* and (ii) have the property that $\mathbb{E}[XY] = 0, \mathbb{E}[X] = 0, \mathbb{E}[Y] = 0$. Again, specify the joint distribution of X and Y.