Spring 2021 18.600 Final Exam: 100 points

1. (10 points) An online class contains 5 first-years, 5 sophomores, 5 juniors and 5 seniors. An instructor randomly divides the students into 4 numbered breakout rooms of 5 students each (with all such divisions being equally likely). Let S_i be the number of seniors in the ith breakout room. Alice and Bob are both seniors in the class.

(a) Compute the probability that Alice is in room 1 and Bob is in room 2. **ANSWER:** Let A be event Alice is in room 1 and B the event that Bob is in room 2. Then $P(AB) = P(A)P(B|A)$. We know $P(A) = 1/4$ (Alice equally likely to be in any room). *Given* that Alice is in room 1, there are 5 slots in room 2 and 19 remaining slots total: Bob is equally likely to be in any of those slots, so we find $P(B|A) = 5/19$ and $P(AB) = 5/76$. As an alternative approach, one can imagine that each room has five slots: there are $20 \cdot 19$ ways to assign Alice to one slot and Bob to another, and 25 of these correspond to Alice in room 1 and Bob in room 2 — so the answer is again $25/(20 \cdot 19) = 5/76$.

(b) Compute $\mathrm{Cov}(S_1, S_2) = E(S_1S_2) - E(S_1)E(S_2)$. **ANSWER:** Let M_i be 1 if ith senior is in room 1, 0 else. Define N_i similarly but with room 2. Then $S_1 = \sum_{i=1}^5 M_i$, $\quad S_2 = \sum_{j=1}^5 N_j$, $E(S_1S_2) = E\big[\big(\sum_{i=1}^5 M_i\big)\big(\sum_{j=1}^5 N_j\big)\big] = \sum_{i=1}^5 \sum_{j=1}^5 E[M_iN_j] = \sum_{i=1}^5 \sum_{j=1}^5 P(M_iN_j = 1)$. The latter probability is zero if $i = j$, but if $i \neq j$ it is the same as what was computed in (a). So we find $E(S_1S_2) = 20 \cdot 5/76 = 25/19$. Also $E(S_1) = \sum_{i=1}^5 E[M_i] = \frac{5}{4} = E(S_2)$. So $\mathrm{Cov}(S_1, S_2) = \frac{25}{19} - \frac{25}{16}$. It makes sense that this quantity is less than zero (so S_1 and S_2 are negatively correlated) since more seniors in room 1 means fewer left for other rooms.

(c) Write $A = S_1 + S_2$ and compute the conditional expectation $E[S_1|A]$ as a function of A. **ANS:** $E[S_1 + S_2|A] = E[A|A] = A$ and $E[S_1 + S_2|A] = E[S_1|A] + E[S_2|A] = 2E[S_1|A]$. Setting RHSs equal and solving gives $E[S_1|A] = A/2$.

(d) Compute the correlation coefficient $\rho(S_1 + S_2, S_3 + S_4)$. **ANS:** $\rho(A, 5 - A) = \rho(A, -A) = -1$.

2. (10 points) Each customer at Alice's ice cream cone stand independently orders one scoop with probability 1/3, two scoops with probability 1/3 and three scoops with probability 1/3. For each $j \geq 1$ let X_j be the number of scoops the jth customer orders. Write $S = \sum_{i=1}^{150} X_i$ for the total number of scoops ordered by the first 150 customers. Compute the following:

(a) The characteristic function $\phi_{X_1}(t)$. **ANSWER:** $E[e^{itX_1}] = \frac{1}{3}(e^{it} + e^{2it} + e^{3it})$.

(b) The characteristic function $\phi_S(t)$. **ANSWER:** $\big(\phi_{X_1}(t)\big)^{150} = \big(\frac{1}{3}(e^{it} + e^{2it} + e^{3it})\big)^{150}$

(c) The variance $\mathrm{Var}(S)$. **ANSWER:** $\mathrm{Var}(X_1) = 2/3$ and $\mathrm{Var}(S) = 150 \cdot 2/3 = 100$ and $\mathrm{SD}(S) = 10$.

(d) Use the central limit theorem to approximation the probability $P(S \leq 320)$. You may use the function $\Phi(a) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ in your answer. **ANSWER:** This is asking for the probability that S is less than or equal to 2 standard deviations above its mean, which is roughly $\Phi(2)$.

3. (10 points) A persistent rock climber is climbing a difficult wall in a rock-climbing gym. The wall has four hand-holds of increasing heights. At each time the climber is in one of five states:

1. State 0: On the mat below the climbing wall.
2. State 1: Holding onto the first hold of the climbing wall.
3. State 2: Holding onto the second hold.
4. State 3: Holding onto the third hold.
5. State 4: At the top celebrating.

Every five seconds (with precise regularity) the climber transitions from one state to another according to the following rules. Whenever the climber is in State 0, the climber transitions to State 1. Whenever the climber is in state i for $i \in \{1, 2, 3\}$ the climber transitions to State $i + 1$ with probability 1/3 and back to State 0 with probability 2/3. Whenever the climber is in State 4, the climber transitions back to State 0.

(a) Write the Markov chain transition matrix corresponding to the climber's state evolution. **ANSWER:** We convert the above into a matrix, where ith row gives the probability of transitioning to any $j \in \{0, 1, 2, 3, 4\}$ if one starts at i.

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 & 0 \\ 2/3 & 0 & 0 & 0 & 1/3 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(b) If the climber is in State 0 at time 0, what is the probability the climber will be back in State 0 after four transition steps? **ANSWER:** By inspection, only the state sequences $0 \to 1 \to 2 \to 3 \to 0$ or $0 \to 1 \to 0 \to 1 \to 0$ are non-zero-chance ways to return to 0 after 4 steps. Add their probabilities: $1 \cdot (1/3)^2(2/3) + 1 \cdot (2/3) \cdot (2/3) = 14/27$.

(c) Over the long term, what fraction of time does the climber spend in each of the 5 states? **ANSWER:**

$$\begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 & 0 \\ 2/3 & 0 & 0 & 0 & 1/3 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix}$$

Multiplying through (starting with second column) we find $\pi_0 = \pi_1$ and $\pi_1/3 = \pi_2$ and $\pi_2/3 = \pi_3$ and $\pi_3/3 = \pi_4$. This implies the vector has the form

$$\begin{pmatrix} \pi_0 & \pi_0 & \pi_0/3 & \pi_0/9 & \pi_0/27 \end{pmatrix}$$

which is a constant multiple of

$$\begin{pmatrix} 27 & 27 & 9 & 3 & 1 \end{pmatrix}$$

Since $\sum_{i=0}^{4} \pi_i = 1$ (and the above entries sum to 67) the vector must take the form

$$\begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} = \begin{pmatrix} 27/67 & 27/67 & 9/67 & 3/67 & 1/67 \end{pmatrix}$$

4. (10 points) On Open-Minded Planet, there are 10^6 people, each of whom holds *either* Opinion A (which happens to be wrong) or Opinion B (which happens to be correct). At time zero 13 unlucky people have acquired Opinion A but 999,987 have Opinion B. During each subsequent time unit, two people (chosen uniformly at random) have a discussion. If they have the same opinion, nothing changes. If they disagree, they talk until one convinces the other—and both parties end up adopting Opinion A with probability 1/2 and Opinion B with probability 1/2 (where this "fair coin toss" is independent of prior tosses). Let X_n be the number of people holding Opinion A after n conversations. One can show (no need to prove) that with probability one X_n *eventually* reaches 0 or 10^6 at which point everyone agrees.

(a) Is the sequence X_n a martingale? Why or why not? **ANSWER:** Yes. Given everything known at stage n, the expected change over the step is always zero, since the probability of somebody switching from A to B always equals the probability of somebody switching from B to A.

(b) What is the probability X_n eventually reaches 10^6 (so a wrong opinion becomes consensus)? **ANSWER:** If T is the time at which consensus is reached then the optional stopping theorem gives $13 = X_0 = E[X_T] = P(X_T = 10^6) \cdot 10^6 + P(X_T = 0) \cdot 0 = P(X_T = 10^6) \cdot 10^6$. Hence $P(X_T = 10^6) = 13/10^6$, a very small number.

Slightly Biased Planet is the same as above except that when two people of opposite opinions argue, they end up both adopting Opinion A with probability $.52 = 13/25$ and Opinion B with probability $.48 = 12/25$. Some pervasive bias makes the wrong opinion *slightly* easier to argue. Let X_n denote the total number of people holding Opinion A after n conversations (again $X_0 = 13$).

(c) Is there a constant $c \in (0, 1)$ for which $M_n := c^{X_n}$ is a martingale? If so, what is c? **ANSWER:** Yes. When somebody switches from B to A, the value M_n gets multiplied by a factor of c. When somebody switches from A to B, M_n gets multiplied by c^{-1}. To keep the expected factor equal to 1, we require $.52c + .48c^{-1} = 1$ or equivalently $.52c^2 + .48 = c$. Solving with the quadratic formula yields $c = 12/13$.

(d) What is the probability that X_n eventually reaches 10^6? *Numerically estimate* this probability by using the approximations $(1-\frac{1}{13})^{13} \approx .36 \approx e^{-1}$ and $(1-\frac{1}{13})^{10^6} \approx 0$. **ANSWER:** Pretend $M_0 = .36$ and either $M_T = 1$ (if $X_T = 0$) or $M_T = 0$ (if $X_T = 10^6$). Optional stopping gives $.36 = E[M_T] = P(M_T = 1) \cdot 1 + P(M_T = 0) \cdot 0 = P(M_T = 1)$. Hence $P(X_T = 10^6) = P(M_T = 0) = .64$. Even though the "bias toward wrongness" is small within each individual argument, the cumulative effect of that bias is that (with 64 percent probability) everyone is wrong in the end.

5. (10 points) Let X_1, X_2, X_3, X_4 be independent exponential random variables with parameter $\lambda = 1$. Write $M = \max\{X_1, X_2, X_3, X_4\}$ and $W = \min\{X_1, X_2, X_3, X_4\}$.

(a) Compute the expectation and variance of W. **ANSWER:** The minimum of four exponentials is again exponential, with parameter given by the sum of the individual parameters — in this case 4. The mean is $1/4$ and variance $1/16$.

(b) Compute the expectation and variance of M. **ANSWER:** This is the radioactive decay problem. The time until the first event is exponential with parameter 4, the time until the next is exponential with parameter 3, etc. Since expectation and variance are both additive (for independent random variables that describe the successive increments) the expectation is $1/4 + 1/3 + 1/2 + 1$ and the variance is $1/4^2 + 1/3^2 + 1/2^2 + 1$.

(c) Compute the expectation $E[X_1 X_2^2 X_3^3 X_4^4]$. **ANSWER:** By independence this is $E[X_1]E[X_2^2]E[X_3^3]E[X_4^4] = 1!2!3!4!$.

6. (10 points) Suppose that (X, Y) has probability density function $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{(-x^2-y^2)/2}$.

(a) Compute the conditional probability $P(X > Y | X > 0)$. **ANSWER:** We want $P(X > Y, X > 0)/P(X > 0)$. The denominator is clearly $1/2$ and since $f_{X,Y}(x, y)$ is rotationally invariant (and the numerator is the probability of a wedge of angle width $3\pi/4$) the numerator is $3/8$. Hence the answer is $3/4$.

(b) Compute the probability density function of $Z = X + 2Y$. **ANSWER:** This is a normal random variable with mean zero and variance $1 + 2^2 = 5$. So answer is $f_Z(z) = \frac{1}{\sqrt{5}\sqrt{2\pi}} e^{-\frac{x^2}{2 \cdot 5}}$.

(c) Compute the expectation $E[X^3Y^3 + X^2Y^2 + XY]$. Given an explicit numerical answer. **ANSWER:** Note $E[X^3] = E[Y^3] = E[X] = E[Y] = 0$ and $E[X^2] = E[Y^2] = 1$. Additivity of expectation and independence gives the answer $0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 = 1$.

7. (10 points) In a city with two political tribes, one tribe expresses genuine frustration with the other by posting 10 articles a day in its news outlet. These articles are surprisingly similar from day to day. Each article is based on one of 6 standard "templates." If $X = (X_1, X_2, \dots, X_{10})$ is the template sequence corresponding to a day's articles, then each X_i is independently equal to

I. ***"They treat us unfairly"*** with probability $1/4$,

II. ***"They criticize us hypocritically"*** with probability $1/8$,

III. ***"They have dark secret motives"*** with probability $1/8$,

IV. ***"We stand up to them"*** with probability $1/4$,

V. ***"We mock them"*** with probability $1/8$,

VI. ***"We are smarter and nobler than they are"*** with probability $1/8$.

(a) Alice does not have time to read all 10 articles, so she programs an AI to categorize the articles and send her just the sequence X. She wants to quantify how many bits of information she learns when she reads X. Compute the entropy $H(X_1)$ and $H(X)$. **ANSWER:** To get $H(X_1)$ we just need to compute $\sum -p_i \log p_i$ where p_i ranges over the 6 given values. We find $2 \cdot (1/4) \cdot 2 + 4 \cdot (1/8) \cdot 3 = 2.5$. Since the X_i are independent we have $H(X) = 10H(X_1) = 25$.

(b) Suppose Alice wants to figure out the value of X_1 by asking a series of yes or no questions. Describe a strategy that minimizes the expected number of questions she needs to ask. How many questions does she expect to ask when she uses this strategy? **ANSWER:** She wants to divide the probability space evenly in two with each question. There are various ways to do this. For example, she could first ask "Is it I, II or III" to determine whether it belongs to the "They" triple or the "We triple." She then asks "Is it the first member of the triple?" If yes, she is done (with 2 questions) and if no she needs to ask "Is it the second member of the triple?" and then she is done. Half the time she finishes in 2 questions, half in 3, so her expected number is 2.5, which is the entropy $H(X_1)$.

(c) Let K be the number of times Template V appears in the sequence X. Compute the entropy $H(K)$. You can leave the answer as an unsimplified sum. **ANSWER:** $\sum_{k=0}^{10} -P(K = k) \log P(K = k)$ where $P(K = k) = \binom{10}{k}(1/8)^k(7/8)^{10-k}$.

8. (10 points) Suppose 10 students take a physical fitness exam. Each student's score on the exam is an independent uniform random variable on $[0, 1]$. Denote the scores by $X_1, X_2, \dots, X_{10}$. Denote by H the highest score and by T the third highest score.

(a) Compute the probability density function $f_H(x)$. **ANSWER:** $F_H(a) = P(\max X_i \le a) = a^{10}$ so $f_H(x) = 10a^9$ on $[0,1]$.

(b) Compute the probability density function $f_T(x)$. **ANSWER:** This is a β random variable with parameters $a-1=7$ (seven values below the third highest) and $b-1=2$ (two values above). This takes the form $x^7(1-x)^2/B(3,8)$. Recall (consult story sheet) that $B(a,b) = (a-1)!(b-1)!/(a+b-1)!$ which in this case comes out to $1/360$.

(c) Compute the conditional probability that the last five scores $X_6, X_7, \ldots, X_{10}$ are in increasing order (i.e., $X_6 < X_7 < \ldots < X_{10}$) *given* that the first six scores are in increasing order (i.e., $X_1 < X_2 < \ldots < X_6$). **ANSWER:** This is the probability all X_j are in increasing order divided by the probability the first six are: so $(1/10)!/(1/6!) = 6!/10!$.

(d) Compute the correlation coefficient $\rho\big(\sum_{j=1}^{8} X_j, \sum_{k=3}^{10} X_k\big)$. **ANSWER:** Generally $\rho(A,B) = \frac{\mathrm{Cov}(A,B)}{\sqrt{\mathrm{Var}(A)\mathrm{Var}(B)}}$. In this example (using bilinearity of covariance and the fact that independent random variables have covariance zero) the numerator is $\sum_{j=3}^{8}\mathrm{Cov}(X_j,X_j) = 6\mathrm{Var}(X_1)$ and the denominator is $8\mathrm{Var}(X_1)$ so the answer is $3/4$.

9. (10 points) Let $X_1, X_2, \ldots, X_5$ be i.i.d. random variables, each with probability density function $\frac{1}{\pi(1+x^2)}$.

(a) Compute the probability that all five X_i lie within the interval $[0,1]$. Give an explicit number. (Recall spinning flashlight story.) **ANSWER:** Each X_i has a $1/4$ chance to be in $[0,1]$ (which corresponds to a $\pi/4$ range of flashlight angles) so answer is $(1/4)^5 = 1/1024$.

(b) Compute the probability density function for $\frac{1}{3}(X_1 + X_2 + X_3)$. **ANSWER:** This has the same law as a Cauchy random variable, so answer is $\frac{1}{\pi(1+x^2)}$

(c) Compute the probability density function for $Y = X_1 + X_2 + X_3$. **ANSWER:** This has the same law as 3 times a Cauchy random variable. Recall generally that when $a > 0$ we have $f_{aX}(x) = a^{-1} f_X(X/a)$. In this case we apply this with $a = 3$ to get $3^{-1}\frac{1}{\pi(1+(x/3)^2)}$

(d) Compute the probability density function for $Z = X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5$. **ANSWER:** Generally jX_j is equivalent in law to the sum of j independent Cauchy random variables. So this overall sum is equivalent in law to the sum of 15 independent Cauchy random variables, which is in turn equivalent in law to $15X_1$. The density function thus takes the form $15^{-1}\frac{1}{\pi(1+(x/15)^2)}$.

10. (10 points) A bag of microwave popcorn contains 500 kernels. For each kernel, suppose the length of time (in seconds) until the kernel pops is an independent gamma random variable with parameter $n = 150$ and $\lambda = 1$. Imagine the kernels are numbered from 1 to 500 and let X_i be the amount of time until the ith kernel pops (so the X_i are i.i.d.).

(a) Compute the probability density function for the "pop time average" $A := \frac{1}{500}\sum_{i=1}^{500} X_i$.
ANSWER: Each X_i be viewed as the sum of 150 i.i.d. exponential random variables each with parameter 1 (by general property of gamma random variables). So $\sum_{i=1}^{500} X_i$ is the sum of 75000 i.i.d. exponential random variables with rate 1, and hence has the density function $e^{-x}x^{74999}/74999!$. By argument similar to that in Problem 9c, we find the answer is $500e^{-500x}(500x)^{74999}/74999!$

(b) Write $G(a) = F_{X_1}(a) = P(X_1 \leq a)$. (You don't have to compute this explicitly.) In terms of the function G, compute the probability density function for the time T at which the last popcorn kernel pops. **ANSWER:** The probability the last kernel pops before time a is $F_T(a) = G(a)^{500}$. Hence $f_T(a) = \frac{\partial}{\partial a} G(a)^{500} = 500G(a)^{499}G'(a)$.

(c) In terms of G, compute the density function for the time at which the first kernel pops.
ANSWER: Call this first popping time U. Then $P(U > a) = (1 - G(a))^{500}$. Hence $F_U(a) = 1 - (1 - G(a))^{500}$ and $f_U(a) = -\frac{\partial}{\partial a}(1 - G(a))^{500} = 500(1 - G(a))^{499}G'(a)$.