Fall 2019 18.600 Final Exam: 100 points
Carefully and clearly show your work on each problem (without writing anything that is technically not true) and put a box around each of your final computations.

NAME: ______________________________

1. (10 points) A school has 60 students (30 boys and 30 girls) and these students are randomly divided into Classroom 1 and Classroom 2, with thirty students assigned to each class. Assume that each of the $\binom{60}{30}$ possible divisions is equally likely. A parent asks the principal, "Does Classroom 1 have at least 29 girls?" and is told that (surprisingly) the answer is yes. The following conversation ensues:

> **Parent zero:** Given what we have just learned, I wonder how likely it is that *all 30 students* in Classroom 1 are girls.
>
> **Parent one:** Well, any given child has a $1/2$ chance to be a girl. So I'd say that if 29 are girls, the chance that the 30th is also a girl is $1/2$.
>
> **Parent two:** No, that's silly. If we know the class has 29 girls, then there is only one girl left among the remaining 31 students, so the conditional probability is $1/31$.
>
> **Parent three:** You are both formulating this the wrong way. The actual conditional probability is even lower than that.

You are summoned to resolve the dispute.

(a) Let A be the event that all 30 students in Classroom 1 are girls. Let B be the event that Classroom 1 has at least 29 girls. Compute the quantities $P(A)$ and $P(B)$.

(b) Now compute the conditional probability $P(A|B)$. Based on this calculation, which (if any) of parents one, two and three is correct?

(c) Let G be the number of girls in Classroom 1 and B the number of boys in Classroom 1. Compute the expectation $E[GB]$.

2. (10 points) A three-year-old child is vying for acceptance at two selective preschools. At each school, acceptance is determined by an entrance exam that measures important preschool skills but is not very reliable. The child's scores on the two exams take the form $S_1 = A + 2B_1$ and $S_2 = A + 2B_2$ where A, B_1, and B_2 are independent normal random variables, each with mean zero and variance one. Informally, A is the student's "entrance exam ability" while B_1 and B_2 are independent noise terms (encoding chance fluctuations).

(a) Compute the expectation $E[S_1 S_2]$.

(b) Compute $\text{Var}(S_1)$ and $\text{Var}(S_2)$ and the correlation coefficient $\rho(S_1, S_2)$.

(c) Compute the conditional expectation $E[S_2|S_1]$ in terms of S_1. That is, express the random variable $E[S_2|S_1]$ as a function of the random variable S_1. (Hint: if it helps, you can argue that $2B_1$ agrees in law with $\sum_{i=1}^4 Y_i$ where the Y_i are independent normal random variables, each with mean zero and variance one.)

3. Jill is an enthusiastic guitarist. Every song she plays has exactly 16 measures. During each measure, independently of all others, she randomly plays of one the five chords she knows:

1. **A major** with probability 1/2
2. **F# minor** with probability 1/8
3. **D major** with probability 1/8
4. **B minor** with probability 1/8
5. **E major** with probability 1/8

Let $X = (X_1, X_2, \ldots, X_{16})$ be the sequence of chords associated to one of Jill's songs.

(a) Compute the entropy $H(X_1)$, i.e., the entropy involved in choosing the first chord.

(b) Compute the entropy $H(X)$, i.e., the entropy involved in choosing an entire song.

(c) Describe a strategy for asking yes-or-no questions to determine X, such that the expected number of questions asked is as small as possible. What is the expected number of questions in this case?

(d) Let $S = \{i : X_i = \text{A major}\}$. In other words, S is the collection of times at which an A major chord is played. Compute $H(S)$ and $H(S, X)$ and $H_S(X)$.

4. A certain small technical university has only five majors: Humanities, Science, Engineering, Business, and Math. During each given week, each student at this university is assigned to exactly one of the five majors. The students at this university change their majors frequently, but they tend to stay in Math a little longer than they do in the other majors. Here is how that works:

> If a student is majoring in Math one week, then the next week she stays in Math with probability 1/2 and transitions to each of the other majors with probability 1/8.
>
> If a student is majoring in any major other than Math one week, then the next week she stays in that major with probability 1/5 and also transitions to each of the other majors with probability 1/5.

Now answer the following:

(a) Represent the major transition process as a Markov chain and write out the five-by-five transition matrix.

(b) Suppose a student is majoring in Math during the first week; compute the probability that he or she will be majoring in Business two weeks later.

(c) Over the long haul, what fraction of the time does a student spend in each of the five majors?

5. (10 points) Let X be an exponential random variable with parameter $\lambda = 1$. For each real number K write $C(K) = E[\max\{X - K, 0\}]$.

(a) Compute $C(K)$ as a function of K for $K \geq 0$.

(b) Compute the derivatives C' and C'' on $[0, \infty)$.

(c) Compute the expectation $E[X^3 + 3X^2 + 3X + 1]$.

6. (10 points) In a certain political party, there are 100 voters actively involved in the process of selecting a nominee. The are 23 candidates, which we number from 1 to 23. At any given time, each voter supports exactly one of these candidates. At the initial time $t = 0$

1. Candidates 1-15 are "lower tier candidates." Each has the support of exactly 2 voters.
2. Candidate 16-20 are "mid-tier candidates." Each has the support of exactly 3 voters.
3. Candidate 21 has the support of 11 voters.
4. Candidate 22 has the support of 21 voters
5. Candidate 23 has the support of 23 voters.

During each unit interval of time (after time $t = 0$) two voters are chosen (uniformly at random from the set of all $\binom{100}{2}$ possible pairs) to have a discussion with each other. If the two voters support the same candidate before the discussion, then the discussion changes nothing; but if they support *different* candidates then *one* of the two voters (chosen by a fair coin toss) switches support to the *other* voter's candidate. This continues until the time T at which all of the voters support the same candidate, at which point that candidate is declared the nominee. (This happens eventually with probability 1, but you don't have to prove that.)

(a) Let $A_i(t)$ be the number of voters supporting candidate i at time t. Is it true that for each $i \in \{1, 2, \ldots, 23\}$, the quantity $A_i(t)$ is a martingale? Explain why or why not.

(b) Compute the probability that for some t with $0 < t < T$ we have $A_{23}(t) = 1$ (so that the initially-leading candidate has lost the support of all but one voter).

(c) Let N be the number of candidates who at *some* point before time T only have the support of a single voter. That is, N is the number of i values for which there is some t such that $0 < t < T$ and $A_i(t) = 1$. Compute the expectation $E[N]$. (Hint: you can use the fact that $\sum_{i=1}^{23} A_i(0) = 100$ and hence $\sum_{i=1}^{23} \big(100 - A_i(0)\big) = 2200$.)

(d) Compute the probability that the candidate who wins the election is somebody who at some time $t > 0$ only had the support of a single voter. (This scenario is called an *epic comeback*.)

7. (10 points) Suppose that the pair (X, Y) is uniformly distributed on the circle $\{(x, y) : x^2 + y^2 \leq 1\}$.

(a) Compute the joint probability density $f_{X,Y}(x, y)$ and the marginal density function $f_X(x)$.

(b) Compute the probability $P(0 < X < Y)$.

(c) Write $Z = X^2 + Y^2$ and work out the density function f_Z.

8. (10 points) Sally the Spammer sends millions of emails every day encouraging the recipients (using various rationales) to grant her unrestricted access to their bank accounts. The times $X_1, X_2, X_3, \ldots$ (measured in years from some initial time) at which Sally is *granted* access to such a bank account form a Poisson point process with parameter $\lambda = 3$. So on average three people per year give Sally access to their bank accounts.

(a) Write $Y = X_3$ and compute the density function f_Y.

(b) What is the probability that Sally is granted bank acount access at least three times during her first year of operation?

(c) Write $Y_0 = 0$ and $Y_k = X_k - \frac{k}{3}$. Is the sequence $Y_0, Y_1, \ldots$ a martingale? Why or why not?

9. (10 points) There are 3 people, each of whom has 4 hats. All 12 hats are tossed into a bin and random divided evenly among the 3 people (so each person gets 4 hats back, with all ways of doing this being equally likely). Let A_i be the event that the ith person gets *all four* of his or her own hats back. Let N be the number of people who get *all four* of their own hats back.

(a) Compute the quantities $a = P(A_1)$ and $b = P(A_1 A_2)$.

(b) Compute $E[N]$ and $\mathrm{Var}(N)$. (You can use the a and b from the previous part in your answer, if that helps.)

(c) Compute $P(N > 0)$. (Again, you can use a and b in your answer, if that helps.)

10. (10 points) Suppose that $X_1, X_2, \dots$ are i.i.d. random variables, each equal to 0 with probability 1/8, 1 with probability 3/4 and 2 with probabiliy 1/8. Write $S_n = \sum_{i=1}^{n} X_i$ and $A_n = S_n/n$.

(a) Compute the moment generating functions $M_{S_{50}}(t)$ and $M_{A_{50}}(t)$.

(b) Compute $E[X_1]$ and $\mathrm{Var}(X_1)$.

(c) Use the central limit theorem to approximate $P(S_{100} > 110)$. You may use the function $\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ in your answer.

(d) Compute the correlation coefficient $\rho(S_{50}, S_{200})$.