18.600 Midterm 2, Fall 2025: 50 minutes, 100 points

1. (15 points) You have hired two promising employees, Avi and Ravi, to sell financial products at your high end investment bank. Let A (resp. R) be the total revenue (in dollars) that Avi (resp. Ravi) generates in sales next year. This revenue depends on many unknown-to-you random factors. Assume that A can be modeled as $A = A_1 + A_2 + \ldots + A_8 + X_1 + X_2 \ldots + X_{12}$ where each A_i depend on attributes and actions of Avi personally (charisma, focus, knowledge, effort, etc.) and each X_i depends on the broader business environment (interest rates, stock prices, etc.) Similarly, assume that R is given by $R = R_1 + R_2 + \ldots + R_8 + X_1 + X_2 + \ldots + X_{12}$ where the R_i depend on Ravi personally and the X_i are the same as for Avi. Assume that all the random variables (the A_i, the R_i and the X_i) are i.i.d. with mean 1 million and standard deviation 1 million.

(a) Compute the correlation coefficient of A and R. **ANSWER:** For each i we find $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i) = 10^{12}$ and similarly for A_i and R_i. Also the covariance of two independent random variables is zero. Using these facts and bilinearity of covariance, we find that $\mathrm{Cov}(A, R) = \sum_{k=1}^{12} \mathrm{Cov}(X_i, X_i) = 12 \cdot 10^{12}$. Using a similar argument (or using additivity of variance for independent random variables) we find $\mathrm{Var}(A) = \mathrm{Var}(R) = 20 \cdot 10^{12}$. Thus $\rho(A, R) = \frac{12 \cdot 10^{12}}{\sqrt{20 \cdot 10^{12} \cdot 20 \cdot 10^{12}}} = 12/20 = 3/5$.

(b) Compute the conditional expectation $E[A|R]$ as a function of R. **ANSWER:** Observe that $R = E[R|R] = E[R_1 + R_2 + \ldots + R_8 + X_1 + X_2 + \ldots + X_{12}|R]$. By additivity of conditional expectation (given R) and symmetry we find that the $E[X_i|R]$ terms are each equal to $R/20$. Since the A_i are independent of R, we have $E[A_i|R] = E[A_i] = 10^6$. So

$$E[A|R] = E[\sum_{i=1}^{8} A_i + \sum_{i=1}^{12} X_i|R] = 8 \cdot 10^6 + \frac{3}{5}R.$$

(c) Compute the expectation and standard deviation of $A + R$. **ANSWER:** $E[A + R] = 40$ million by additivity of expectation. And

$$\mathrm{Var}(A + R) = \mathrm{Var}\Big(\sum_{i=1}^{8} A_i + \sum_{i=1}^{8} R_i + 2\sum_{i=1}^{12} X_i\Big) = 8 \cdot 10^{12} + 8 \cdot 10^{12} + 12 \cdot 4 \cdot 10^{12} = 64 \cdot 10^{12},$$

so $\mathrm{SD}(A + R) = 8$ million.

2. (20 points) Naomi is a New Jersey grandmother whose yard is heavily frequented by deer in the winter. To discourage deer from eating her plants, Naomi (who is also a retired baseball pitcher) resolves to sit on her back porch next to a bucket of snowballs and hurl one at each deer who enters her yard. Assume that the deer come one at a time, and the arrival times $D_1, D_2, \ldots$ form a Poisson point process, with two deer expected to arrive each hour. Assume that each time a deer arrives, Naomi successfully hits it with a snowball with probability $1/3$ (independently of everything else).

(a) Compute the mean and variance of the total number of deer to arrive during the first 3 hours. **ANSWER:** Number is Poisson with parameter $\lambda = 3 \cdot 2 = 6$. So mean and variance are both 6.

(b) Compute the mean and variance of the total number of deer successfully hit by snowballs during the first 3 hours. **ANSWER:** Each deer is "counted" independently with probability $1/3$, and pset problem shows number counted is Poisson with parameter $6 \cdot (1/3) = 2$. So mean and variance are both 2.

(c) Naomi has only 20 snowballs in her bucket, so she will go back in the house after throwing a snowball at the 20th deer at time D_{20}. Compute the probability density function for D_{20}. **ANSWER:** This is a Gamma with parameters 2 and 20. $f(x) = 2 \cdot (2x)^{19} e^{-2x}/19!$

(d) Compute the probability that Naomi hits exactly ten deer before she goes back inside. **ANSWER:** Number hit binomial with $p = 1/3$, $n = 20$, so answer is $\binom{20}{10}(1/3)^{10}(2/3)^{10}$.

3. (20 points) There are 100 students in a first-year writing course. Each student writes one essay. For each $i \in \{1, 2, \dots, 100\}$ the ith student's essay is given a score X_i that indicates its "insightfulness" and a second score Y_i that indicates its "persuasiveness." Assume that the scores $X_1, X_2, \dots, X_{100}$ and $Y_1, Y_2, \dots, Y_{100}$ are all independent random variables. The probability density function for each of the X_i is given by $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. The probability density function for each of the Y_i is given by $g(x) = \frac{1}{\pi(1+x^2)}$. Let $X = \sum_{i=1}^{100} X_i$ be the class's "total insightfulness score" and $Y = \sum_{i=1}^{100} Y_i$ the class's "total persuasiveness score." Let Z be the number of students rated "more-persuasive-than-insightful" — i.e., the number of values $i \in \{1, 2, \dots, 100\}$ for which $X_i < Y_i$. If it helps, you may use the function $\Phi(a) := \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$ in your answers below.

(a) Compute the probability $P(-100 < X < 100)$. **ANSWER:** Since X is normal with mean 0 and variance 100 (and SD 10), answer is probability that normal random variable is between -10 and 10 SDs above mean, which is $\Phi(10) - \Phi(-10)$ (extremely close to 1).

(b) Compute the probability $P(-100 < Y < 100)$. **ANSWER:** This is same as $P(-1 < Y/100 < 1)$. Since $Y/100$ is average of i.i.d. Cauchy random variables, it is itself Cauchy, and spinning flashlight story gives answer 1/2.

(c) Compute the mean and standard deviation of Z. **ANSWER:** By symmetry, $P(X_1 < Y_1) = P(-X_1 < -Y_1) = P(X_1 > Y_1) = 1/2$. So Z is binomial random variable with $n = 100$ and $p = 1/2$. Hence $E[Z] = np = 50$ and $\text{Var}(Z) = npq = 25$ and $\text{SD}(Z) = 5$.

(d) Give an approximation for the probability $P(Z \geq 40)$. **ANSWER:** 40 is two SDs below mean, so answer is roughly $1 - \Phi(-2) = \Phi(2) \approx .977$.

4. (15 points) **The Semicircle** is the name of an enormous concert venue in the shape of a semicircle of radius 100 meters—a.k.a. *one hectometer*). Using the hectometer as our distance unit, we can parameterize the semicircle by the set $S = \{(x, y) : x^2 + y^2 \leq 1, x \geq 0\}$. When you purchase a ticket to the upcoming *Bad Bunny* concert at the venue, your seat location is chosen uniformly at random from S. Let (X, Y) be the coordinates of the seat location. Let $R = \sqrt{X^2 + Y^2}$ be the distance of your seat from the origin (where the singer is located). The cost of the ticket in dollars is $C = 1000/R$ (so seats nearer the origin are more expensive).

(a) Compute the conditional expectation $E[X|Y]$ as a function of Y. **ANSWER:** For any $y \in [-1, 1]$ the conditional law of X given $Y = y$ is uniform on $[0, \sqrt{1 - y^2}]$, and hence the conditional expectation is $\frac{1}{2}\sqrt{1 - y^2}$. Conclude that $E[X|Y] = \frac{1}{2}\sqrt{1 - Y^2}$.

(b) The seats seem rather expensive. Compute the expectation of the cost, i.e. $E[C]$. **ANSWER:** Note that S has area $\pi/2$ so $f(x, y) = 2/\pi$ for $(x, y) \in S$. Then

$$E[C] = \frac{2}{\pi}\iint_S f(x,y)\frac{1000}{\sqrt{x^2+y^2}}dxdy = \frac{2}{\pi}\int_0^{\pi}\int_0^1 (1000/r) \cdot rdrd\theta = \frac{2}{\pi}\int_0^{\pi}\int_0^1 1000drd\theta = 2000$$

(c) Actually, you only have \$5000 in your bank account, so there is some danger that the ticket purchase will put you in debt. Compute the probability $P(C \leq 5000)$. **ANSWER:** Write $R = \sqrt{X^2 + Y^2}$. Then $P(C \leq 5000) = P(R \geq 1/5) = 1 - P(R \leq 1/5) = 1 - (1/5)^2 = .96$.

(d) Compute the conditional probability $P(C \leq 5000|C \leq 10000)$. **ANSWER:** By same argument $P(C \leq 1000) = P(R \geq 1/10 = .99$. Then $P(C \leq 5000|C \leq 10000) = .96/.99 = 32/33$.

5. (15 points) A new student-focused restaurant is opening in Kendall Square. Given the high rents and other expenses, and the limits to what students can pay, it is hard to operate a restaurant in a way that everybody will consider "acceptable." In fact, your Bayesian prior for the fraction of people who consider the restaurant acceptable is uniform on $[0, 1]$. In other words, you may assume that there is a uniform random variable $p \in [0, 1]$ that represents the restaurant quality. Once we are *given* the value of p, each customer who visits the restaurant considers it "acceptable" with probability p and "unacceptable" with probability $1 - p$ (independently from one customer to the next). **NOTE:** If it helps, you may use the fact that a Beta (a, b) random variable has expectation $a/(a + b)$ and density $x^{a-1}(1 - x)^{b-1}/B(a, b)$, where $B(a, b) = (a - 1)!(b - 1)!/(a + b - 1)!$.

(a) Suppose three people visit the restaurant one at a time. Compute the probability of the event A that the first and second visitors declare the restaurant acceptable *and* the third visitor declares it unacceptable. **ANSWER:** Recalling Rotten Tomatoes (Pólya's urn) this is $1/2 \cdot 2/3 \cdot 1/4 = 1/12$.

(b) Given that the event A occurs (i.e. that of first three people who visit the restaurant, the first two declare it acceptable and the third declares it unacceptable) what is your conditional probability density function for p? In other words, what is your Bayesian posterior for p after these observations? **ANSWER:** This is Beta with $a - 1 = 2$ and $b - 1 = 1$. So $a = 3$ and $b = 2$ and density is $x^{a-1}(1 - x)^{b-1}/B(a, b) = x^2(1 - x)/(2!/4!) = 12x^{a-1}(1 - x)^{b-1}$.

(c) *Given* these first three reports (two acceptable, one unacceptable) what is the *conditional probability* that the fourth visitor will find the restaurant acceptable? **ANSWER:** 3/5, using Pólya's urn argument as (a).

6. (15 points) 100 people take a new weight loss drug for one week as part of a clinical trial. Each person who takes the drug (independently) loses

0 pounds with probability 1/4,

1 pound with probability 1/2, and

2 pounds with probability 1/4.

For $i \in \{1, 2, \dots, 100\}$ let Z_i be the number of pounds the ith person loses. Let $A = \frac{1}{100}\sum_{i=1}^{100} Z_i$ be the average amount of weight lost.

(a) Compute the moment generating functions $M_{Z_1}(t)$ and $M_A(t)$. Give exact formulas. **ANSWER:** $M_{Z_1}(t) = E[e^{tZ_1}] = \frac{1}{4}e^{0\cdot t} + \frac{1}{2}e^t + \frac{1}{4}e^{2t}$ and $M_A(t) = \big(M_{Z_1}(t/100)\big)^{100} = \Big(\frac{1}{4} + \frac{1}{2}e^{t/100} + \frac{1}{4}e^{2t/100}\Big)^{100}$.

(b) Compute the characteristic function $\phi_A(t)$. Give an exact formula. **ANSWER:** $\phi_A(t) = M_A(it) = \Big(\frac{1}{4} + \frac{1}{2}e^{it/100} + \frac{1}{4}e^{2it/100}\Big)^{100}$