EECS 126 Probability and Random Processes | UC Berkeley, Spring 2023
Thomas Courtade | Febuary 16, 2023

Midterm 1

Last Name	First Name	SID

Left Neighbor First and Last Name	**Right Neighbor** First and Last Name

***Rules*.**

- **Write in your SID on every page to receive 1 point.**
- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- You have 80 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $80 \cdot X\%$ time on completing the exam).
- This exam is closed-book. You may reference one double-sided sheet of notes. No calculator or phones allowed.
- Remember the Berkeley Honor Code: "*As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.*" Any violation of academic integrity will be taken seriously, and could result in disciplinary consequences.

Problem	out of
SID	1
Problem 1	10
Problem 2	15
Problem 3	15
Problem 4	20
Problem 5	20
Problem 6	24
Problem 7	15
Total	120

 Student ID (write on every page for 1 point): ____________

1 Something Good [7 + 3 points]

a) What's something you are looking forward to this semester? (doesn't have to be related to this class)

b) Write down as many passcodes from lecture as you can remember.

2 Monkey Business [5 + 10 points]

A monkey types on a keyboard with 27 keys (corresponding to letters a-z, plus a period '.').

a) Assuming the monkey types each character independently and uniformly at random, what is the probability they type "class." on their first try? (Note the period at the end of the word 'class')

b) Let $\ell(i)$ denote the number of letters in the ith word in the English dictionary (all words are composed of only letters a-z, no periods). For example, if 'aardvark' is the 3rd word in the dictionary, then $\ell(3) = 8$. Assuming there are N words in the dictionary, use the axioms of probability to show that

$$\sum_{i=1}^{N} \left(\frac{1}{27}\right)^{\ell(i)} \leq 27.$$

Hint: Define N disjoint events.

a) Since we need to hit the 6 symbols in order, and keys are struck uniformly, the probability is $(1/27)^6$.

b) Let A_i be the probability the monkey types the ith word in the dictionary, followed by a period, on their first try. Then, $A_1, A_2, \ldots, A_N$ are disjoint events, and $P(A_i) = (1/27)^{\ell(i)+1}$. Hence, by the axioms of probability

$$\sum_{i=1}^{N}(1/27)^{\ell(i)+1} = \sum_{i=1}^{N} P(A_i) = P(\cup_{i=1}^{N} A_i) \leq 1.$$

Rearranging gives the claim.

Concepts tested: Computing simple probabilities from a given model; formulating suitable events; probability axioms.

3 Bear Territory [15 points]

A zoologist observes $B \sim \text{Poisson}(\mu)$ bears living on the prairie. Bear i has a territory with "range parameter" $R_i \sim \mathcal{N}(r, \sigma^2)$. Conditioned on R_i, the territory of bear i has area $X_i \sim \text{Uniform}(R_i^2, 3R_i^2)$, independent of the number of bears B. None of the bear territories overlap. What is the expected total territory area T occupied by all bears on the prairie?

This is like the "random sum of random variables" example we saw in class to illustrate the usefulness of iterated expectation. Since $T = \sum_{i=1}^{B} X_i$, and areas are independent of B, we can compute

$$\mathrm{E}(T) = \mathrm{E}(\mathrm{E}(T \mid B)) = \mathrm{E}(\mathrm{E}(\textstyle\sum_{i=1}^{B} X_i \mid B)) = \mathrm{E}(B\,\mathrm{E}(X_1)) = \mathrm{E}(B)\,\mathrm{E}(X_1).$$

It remains to find the expected area of one bear's territory, which can also be evaluated by iterated expectation, since the distribution of X_i is uniform once we fix R_i:

$$\mathrm{E}(X_1) = \mathrm{E}(\mathrm{E}(X_1 \mid R_1)) = \mathrm{E}(2R_1^2) = 2(\mathrm{var}(R_1) + \mathrm{E}(R_1)^2) = 2(\sigma^2 + r^2).$$

So, the expected total area is

$$\mathrm{E}(T) = 2\mu(\sigma^2 + r^2).$$

Concepts tested: Iterated expectation; linearity of expectation; variance decomposition in terms of second moment.

4 Noisy Observations [10 + 10 points]

Let X and Z be independent standard normals. Suppose we consider

$$Y = \alpha X + Z,$$

where $\alpha \in \mathbb{R}$.

a) Find the conditional density $f_{X|Y}(x|y)$. What is the distribution of X given $\{Y = y\}$? Specify any parameters of the distribution as needed.

b) Compute the conditional expectations $\mathrm{E}[X|Y = y]$ and $\mathrm{E}[Y|X = x]$.

Hint: The density of $\mathcal{N}(\mu, \sigma^2)$ is $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

This is the "bits in noise" example we gave in class, but with X Gaussian instead of ± 1 valued. Just like we did in class, the idea is that we should use Bayes to convert from $f_{Y|X}$ (which is deducible directly from our model) to $f_{X|Y}$.

a) Note that $Y \sim N(0, 1+\alpha^2)$, since sums of independent Gaussians are Gaussian. Also, similar to our "bits in noise" example in class, we have $Y \sim \mathcal{N}(\alpha x, 1)$ conditioned on $\{X = x\}$. By continuous Bayes, we have

$$f_{X|Y}(x|y) = \frac{f_X(x)}{f_Y(y)} f_{Y|X}(y|x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\frac{1}{\sqrt{2\pi(1+\alpha^2)}} e^{-\frac{y^2}{2(1+\alpha^2)}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\alpha x)^2}{2}} = \frac{1}{\sqrt{2\pi \frac{1}{1+\alpha^2}}} e^{-\frac{1}{2\frac{1}{1+\alpha^2}}(x - \frac{\alpha}{1+\alpha^2} y)^2}.$$

We recognize this as a Gaussian density, hence

$$X|\{Y = y\} \sim N\left(\frac{\alpha}{1+\alpha^2} y, \frac{1}{1+\alpha^2}\right).$$

b) By the last part, we have $\mathrm{E}[X|Y = y] = \frac{\alpha}{1+\alpha^2} y$. The other conditional expectation is simply $\mathrm{E}[Y|X = x] = \alpha x$ since $Y \sim \mathcal{N}(\alpha x, 1)$ conditioned on $\{X = x\}$.

Concepts tested: Properties of Gaussians; densities and conditional densities; continuous Bayes; conditional expectation.

5 Lucky Draws [5 + 5 + 10 points]

There are n red balls and n green balls in a bucket. You sequentially draw all $2n$ balls, without replacement from the bucket, and record the sequence of the colors you observe. A draw is considered a "lucky draw" if it is the same color as the previous draw. Let N denote the number of lucky draws after drawing all balls.

a) Formulate a suitable probability space for this problem.

b) Express the random variable N as an explicit function $N(\omega)$ of the underlying sample $\omega \in \Omega$. It is acceptable to leave your answer in the form of a sum.

c) Find $\mathrm{E}[N]$.

This problem emphasizes the formulation of a suitable probability space, and the formulation of random variables as a function on that space. It is a modification of the "hat matching" example we saw in class (in that case, Ω was the set of permutations, P was the uniform measure, and N was the number of matched hats – expressible as a sum of indicators).

a) Let Ω equal the set of binary vectors of length $2n$, each with exactly n ones. Let $\mathcal{F} = 2^\Omega$, and P be the uniform measure; i.e., $P(\omega) = 1/\binom{2n}{n}$.

b) Letting $\omega = (\omega_1, \ldots, \omega_{2n})$, we have

$$N(\omega) = \sum_{i=2}^{2n} 1_{\{\omega_i = \omega_{i-1}\}}(\omega).$$

c) Writing out N explicitly, as we did above, suggests we should consider indicator random variables. For $i = 2, \ldots, 2n$, let X_i denote the indicator that draw i and $i-1$ share the same color. Once we fix the color of draw $i-1$, there are $n-1$ balls of the same color (out of $2n-1$ remaining) to put in position i to make draw i lucky. Hence, $\mathrm{E}[X_i] = P\{X_i = 1\} = \frac{n-1}{2n-1}$, and by linearity of expectation

$$\mathrm{E}[N] = \sum_{i=2}^{2n} \mathrm{E}[X_i] = (n-1).$$

Concepts tested: Formulating probability spaces and random variables; indicator random variables; linearity of expectation.

6 *N*th Times a Charm [8 + 8 + 8 points]

You have a coin that turns heads with probability p and tails with probability $1-p$. For a given integer $k \geq 1$, let N denote the number of independent flips until you see exactly k heads.

a) Find the pmf of N.

b) Compute the moment generating function $M_N(t)$.

c) Compute the variance of N.

Note: You might find the series $\sum_{n\geq 1} n^2 q^n = \frac{q(1+q)}{(1-q)^3}$ (for $|q| < 1$) helpful in part (c).

This problem parallels our discussions of the Binomial distribution in class, which can be seen as a sum of independent Bern(p) random variables. Here, we switch things up slightly and instead consider a sum of independent Geom(p) random variables, and find pmf, mgf and variance.

a) For $\{N = n\}$, we need to have a heads in the nth position, and exactly $k-1$ heads in the previous $n-1$ positions. These events are independent, with the probability of the former equal to p, and the latter equal to $\binom{n-1}{k-1}(1-p)^{n-k}p^{k-1}$. We also must have $n \geq k$. Hence,

$$p_N(n) = \binom{n-1}{k-1}(1-p)^{n-k}p^k, \quad n \geq k.$$

b) Since $N = X_1 + \cdots X_k$, where the X_i's are i.i.d. Geom(p), we have $M_N(t) = M_{X_1}(t)^k$. So, we just need to compute the mgf of a Geom(p) distribution. To that end,

$$M_{X_1}(t) = \mathrm{E}[e^{tX_1}] = \sum_{n\geq 1}(1-p)^{n-1}pe^{tn} = pe^t \sum_{n\geq 0}((1-p)e^t)^n = \frac{pe^t}{1-(1-p)e^t}.$$

(The series converges when $(1-p)e^t < 1$, but you don't have to worry about this). Hence,

$$M_N(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^k, \qquad t < -\log(1-p).$$

c) Since variance is additive on independent summands,

$$\mathrm{Var}(N) = k\,\mathrm{Var}(\mathrm{Geom}(p)) = k\frac{1-p}{p^2}.$$

In case you didn't already know it, the variance of Geom(p) can be obtained from the hint:

$$\mathrm{E}[X_1^2] = \frac{p}{1-p}\sum_{n\geq 1} n^2(1-p)^n = \frac{2-p}{p^2} \quad \Rightarrow \quad \mathrm{Var}(X_1) = \mathrm{E}[X_1^2] - \mathrm{E}[X_1]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

 Student ID (write on every page for 1 point): ____________

Concepts tested: reasoning to find pmfs of discrete distributions; decomposing a "complicated" random variable in terms of a simpler random variables with known distributions; computing mgf and variance.

7 Reasonable Grading [15 points]

A professor claims that they can grade exam problems simply by sniffing the page on which the solution is written. Each problem is graded on a scale from 0-4 points (4 = A, 3 = B, 2 = C, 1 = D, 0 = F). When the professor smells 'A' work, they immediately recognize it as 'A' work; likewise, when the professor smells 'F' work, they also recognize it without mistake. However, when the student solution deserves 1,2 or 3 points, the smells can be hard to discern, and the professor adds 1 point, subtracts 1 point, or gets the score correct, each with probability 1/3, independent of all other problems.

Throughout the semester, 25 exam problems are graded for each student, and the scores are averaged and then rounded to the nearest integer determine the course letter grade (4 = A, 3 = B, etc.).

Suppose that a particular student deserved an average score of 3.1 (i.e., this would have been their score if the professor made no mistakes in grading the questions). Find a reasonably good numerical upper bound on the probability that a student receives a lower grade than the B they deserve.

The model here is basically the same as the example given in class to illustrate WLLN with noisy reader scores. However, instead of letting the number of problems grow to infinity, you are asked to bound the probability for a finite number of samples (like we did in examples using concentration inequalities, including the proof of the WLLN itself).

The professor computes the student score as $3.1 + Z$, where Z is the grading error defined by $Z := \frac{1}{25}\sum_{i=1}^{n} X_i$, with i.i.d. $X_i \sim \text{Unif}(\{-1, 0, +1\})$, and n the number of questions where the student deserved a 1,2 or 3. Hence, the probability a student gets a lower course score than they deserve is

$$P\{3.1 + Z < 2.5\} = P\{Z < -0.6\} = P\{\sum_{i=1}^{n} X_i < -15\} = \frac{1}{2}P\{|\sum_{i=1}^{n} X_i| > 15\},$$

where the last equality is by symmetry of the distribution of Z. Applying Chebyshev gives

$$P\{Z < -0.6\} \leq \frac{1}{2}\frac{n\,\text{Var}(X_1)}{15^2} = \frac{n}{3 \cdot 15^2} \leq \frac{25}{3 \cdot 15^2} = \frac{1}{27} \approx 3.7\%.$$

Just for curiosity's sake, we remark that we could instead apply a Chernoff bound to find the 15x improvement $P\{Z < -0.6\} \leq 0.000679906 < 0.067\%$, but this is not easy to do by hand.

Concepts tested: Using random variables to model a problem given a description; concentration inequalities (Chebyshev, in particular).