American University of Beirut

MATH/STAT 233: Probability

2022–2023 Spring

Midterm 2 (Tuesday, March 21, 17:30 – Nicely 212)

How it works: You receive the questions in advance. At the beginning of the exam, you declare which of the questions you have managed to solve. For each question, one student (from those who have solved it) will be called to show their solution on the board.

What is permitted and what is not permitted: Before the exam, you can (and are encouraged to) discuss the problems with the other students in class but not with external people. Getting help from internet forums is also not permitted. Feel free to use your class notes, books, computers, or Wikipedia.

Problem 1 (A paradox)**.** Consider the following paradoxical situation:

> We have three jars, each containing two balls. Jar #1 has two blue balls, jar #2 has two red balls, and jar #3 has one blue ball and one red ball. We pick a jar at random and draw a ball from it at random without looking. The ball we picked is equally likely to be blue or red. Likewise, the other ball in the jar we picked is equally likely to be blue or red. Therefore, there is a $1/2$ probability that the two balls in the jar we picked have different colors. But this is clearly false, since the probability of picking the jar with different colors is $1/3$.

Find a resolution for this paradox. Explain what is wrong with the above reasoning and provide a correct analysis.

Problem 2 (Expectation of absolute value)**.** In class, we exploited the inequality

$$\mathbb{E}\left[|X|\right] \leq \sqrt{\mathbb{E}[X^2]}$$

and referred to it as (a special case of) Jensen's inequality. Give a direct proof of the above inequality by examining the variance of $|X|$.

Problem 3 (Poisson vs. normal distributions)**.**

(a) Let X and Y be independent Poisson random variables with parameters μ_X and μ_Y, respectively. Prove that $Z := X + Y$ is also a Poisson random variable, and find its parameter.
[*Hint*: First, identify the possible values of Z. For each possible value k, compute $\mathbb{P}(Z = k)$ using the law of total probability, by breaking the possibilities based on the values of X.]

(b) Use induction (or any other approach) to show that the sum of any finite collection of independent Poisson random variables is again a Poisson random variable.

(c) Use the latter result to argue that when μ is large, the Poisson distribution with parameter μ can be approximated with a normal distribution. Formulate a precise version of this claim as a corollary of the central limit theorem.

(d) What is wrong with the following reasoning?

> **Claim**. Every Poisson distribution can be approximated by a normal distribution.
> "*Argument*". Let $\mu > 0$ be arbitrary.
> Let n be very large, and set $\varepsilon := \mu/n$. Consider n independent Poisson random variables $Y_1, Y_2, \ldots, Y_n$, each with parameter ε. By the result of part (b), the sum $X := Y_1 + Y_2 + \cdots + Y_n$ has a Poisson distribution with μ as the parameter. On the other hand, the central limit theorem tells us that the distribution of X is approximately normal. Thus, the Poisson distribution with parameter μ is approximately normal. $\square$

Problem 4 (Powers of an exponential RV). Let T be an exponential random variable with rate 1, and let $\beta > 0$ be a positive constant. Define a new variable $X := T^{1/\beta}$.

(a) Find the probability density function of X.
[*Hint*: For $b \geq a \geq 0$, write $\mathbb{P}(a \leq X \leq b)$ in terms of the density of T, and change the variable of integration to make the integration limits a and b. Later we will learn a more straightforward approach.]

(b) Find $\mathbb{E}[X]$.

(c) Find $\mathbb{V}\mathrm{ar}[X]$.

Problem 5 (Cheating and embarrassment). A group of researchers would like to estimate how widespread cheating is during exams among undergraduate students. One difficulty is that students are often too embarrassed to admit to cheating. To circumvent this difficulty, the researchers come up with the following strategy using a deck of cards. A fraction q_0 of the cards say "Answer *no*", a fraction q_1 of the cards say "Answer *yes*", and a fraction q_2 of the cards say "Answer truthfully". When interviewing a student, the researchers first completely shuffle the deck of cards. They then ask the interviewee to draw a card from the deck at random and follow the instruction on the card when asked the question "Have you ever cheated during an exam?". The point is that the student would not be embarrassed to answer truthfully when instructed because the researchers have no way of knowing if the student had drawn a *yes* card or a *truth* card. Nevertheless, in a large number of interviews, the truthful answers shift the balance of positive responses. The researchers can then use the magnitude of this shift to estimate the true proportion of the students who have cheated in an exam.

Let p denote the true proportion of the students who have cheated during an exam. We assume that the students would follow the instruction of the card accurately.

(a) What is the probability that a random student answers *yes* to the question?

(b) If a random student has answered *yes* to the question, what is the probability that they have indeed cheated?

(c) If a random student has answered *yes* to the question, what is the probability that they have picked a *truth* card?

(d) How should the researchers choose the parameters q_0, q_1 and q_2 in order to make the students more comfortable about following the instruction accurately.

Suppose the researchers interview a random sample of 4000 students. Let R denote the fraction of the interviewed students who answered *yes* to the cheating question. For simplicity, we assume that the sample is taken with replacement.

(e) Based on the observed value of R, what would be your estimate of the proportion p? Would your estimate be reasonable if R turns out to be less than q_1?

(f) Find the expected value and the variance of R in terms of the parameters q_0, q_1, q_2 and the unknown p.

Suppose the researchers have decided to go with the choices $q_0 = q_1 = 2/5$ and $q_2 = 1/5$.

(g) Compute or estimate the probability that your estimate of p becomes larger than 35% if in reality $p = 25\%$.