Math 472: Homework 01

Due Wednesday, Jan 21

Remark: The problems in this homework are intended to either introduce you to `R` or introduce/review some central concepts in probability that we will need throughout this course. For the remainder of this course, any terms defined using ***this format*** are precise definitions and should be memorized precisely (they are fair game for in-class quizzes after the homework is due). Text in **bold** should be definitely reviewed if you are not familiar with it.

Problem 1. Install a bunch of software on your laptop:

(a) Install the statistical software `R`, available at the `https://cran.rstudio.com/`

(b) Install `RStudio Desktop`, available at `https://posit.co/download/rstudio-desktop/`. (If you already know R with a different IDE, you can skip this.)

(c) Do the `R` tutorial. (There are two optional exercises in this, but you don't have to turn anything in.)

Problem 2 (Exercise 1.2 in the textbook)**.** Are some cities more windy than others? Does Chicago deserve to be nicknamed 'The Windy City'? Given below are the average wind speeds (in miles per hour) for 45 selected U.S. cities:

8.9	7.1	9.1	8.8	10.2	12.4	11.8	10.9	12.7
10.3	8.6	10.7	10.3	8.4	7.7	11.3	7.6	9.6
7.8	10.6	9.2	9.1	7.8	5.7	8.3	8.8	9.2
11.5	10.5	8.8	35.1	8.2	9.3	10.5	9.5	6.2
9.0	7.9	9.6	8.8	7.0	8.7	8.8	8.9	9.4

(a) Use R to construct a relative frequency histogram for these data. (Play around with the number of breaks and find one that is reasonable).

(b) The value 35.1 was recorded at Mt. Washington, New Hampshire. Does the geography of that city explain the magnitude of its average wind speed?

(c) The average wind speed for Chicago is 10.3 miles per hour. What percentage of the cities have average wind speeds in excess of Chicago's?

(d) Do you think that Chicago is unusually windy?

Problem 3. A pair of events A and B cannot be simultaneously *mutually exclusive* and *independent*. Prove that if $P(A) > 0$ and $P(B) > 0$, then:

(a) If A and B are mutually exclusive, then they cannot be independent.

(b) If A and B are independent, then they cannot be mutually exclusive.

Problem 4. Let $\mathbb{P}$ be a probability measure and let B be an event with $\mathbb{P}[B] > 0$. Define the function $\mathbb{P}'$ by

$$\mathbb{P}'[A] := \mathbb{P}[A \mid B]$$

for every event A. Prove that $\mathbb{P}'$ is a probability measure.

Problem 5 (Exercise 1.20 in the textbook)**.** Weekly maintenance costs for a factory, recorded over a long period of time and adjusted for inflation, tend to have an approximately normal distribution with an average of \$420 and a standard deviation of \$30. If \$450 is budgeted for next week, what is an approximate probability that this budgeted figure will be exceeded?

Problem 6 (This is essentially Example 2.1 in the textbook)**.** A drone manufacturer has six seemingly-identical drones available for shipping. Unknown to her, two of the six have defective optics. A particular order calls for two drones and is filled by randomly selecting two of the six that are available.

(a) Label four good drones by G_1, G_2, G_3, G_4 and the two defective drones by B_1 and B_2. List the sample space for this experiment.

(b) Let A denote the event that the order is filled with two non-defective drones. List the sample points in A.

(c) Assign probabilities to each of the simple events in such a way that the information about the experiment is used (and such that the probability measure axioms are met).

(d) Find the probability of A.

(e) The following R code

```
drones = c(1,1,1,1,0,0)
sum(sample(drones,2,replace=FALSE))==2
```

represents the drones as 1's (nondefective) and 0's (defective), samples 2 of them without replacement, and returns TRUE iff both drones are nondefective.

Use the R function `replicate()` to repeat this experiment 10,000 times and use your result to estimate the probability of A.

Problem 7 (Exercise 2.73 in the textbook)**.** In 1865, the Austrian monk Gregor Mendel proposed a theory of inheritance which later became the foundation of modern genetics. In his experiments on flower color, Mendel studied heterozygous plants possesing two alleles:

$$\begin{aligned} \text{w} &= \text{ recessive white-color allele} \\ \text{R} &= \text{ dominant red-color allele.} \end{aligned}$$

When two heterozygous plants were crossed, approximately 3/4 of the offspring had red flowers and approximately 1/4 had white flowers. The following table summarizes the possible allele combinations for the offspring; each parent contributes one allele, chosen at random, to the offspring:

	Parent 1	
Parent 2	w	R
w	ww	wR
R	wR	RR

Assume each parent is equally likely to pass on either allele (w or R), and that the offspring has red flowers if and only if at least one of its alleles is dominant (R), the offspring will have red flowers.

(a) What is the probability that an offspring has at least one dominant allele?

(b) What is the probability that an offspring has at least one recessive allele?

(c) Given that an offspring has red flowers, what is the probability that it has exactly one recessive allele?

Problem 8 (Exercise 2.86 in the textbook)**.** Suppose that A and B are two events such that

$$\mathbb{P}[A] = .8 \quad \text{and} \quad \mathbb{P}[B] = .7.$$

(a) Is it possible that $\mathbb{P}[A \cap B] = .1$? Why or why not?

(b) What is the smallest possible value for $\mathbb{P}[A \cap B]$?

(c) Is it possible that $\mathbb{P}[A \cap B] = .77$? Why or why not?

(d) What is the largest possible value for $\mathbb{P}[A \cap B]$?

Problem 9 (The Law of Rare Events)**.** A random variable Y is said to have a ***Poisson probability distribution*** with parameter $\lambda > 0$, written $Y \sim \text{Pois}(\lambda)$ if its probability mass function is

$$\mathbb{P}[Y = y] = \frac{\lambda^y}{y!}e^{-y}, \quad y \in \{0, 1, 2, \ldots\}.$$

A random variable X is said to be ***binomially distributed***, written $X \sim \text{Bin}(n, p)$, with parameters n (called the "size" or "number of trials") and p (the "success probability") if its probability mass function is

$$\mathbb{P}[X = x] = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \ldots, n\}.$$

[These are two important distributions; see Sections 3.4 and 3.8 in the textbook for good discussions of both.] In this problem, we'll show how to approximate a binomial distribution with a Poisson distribution. (This approximation is sometimes called *The Law of Rare Events*).

(a) Show that for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

(b) For fixed $\lambda > 0$ and $k \in \{0, 1, 2, \ldots\}$, show that

$$\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k = \frac{\lambda^k}{k!}.$$

(c) Fix $\lambda > 0$ and $k \in \{0, 1, 2, \ldots\}$. Assume that $p = \lambda/n$ (which is getting smaller as $n \to \infty$). Use parts (a) and (b) to show that

$$\mathbb{P}[Y = k] = \lim_{n\to\infty} \mathbb{P}[X = k]$$

(d) In part (c) you showed that if n is large and p is small (and $\lambda = np$ isn't too big), then a random variable $X \sim \text{Bin}(n, p)$ is well-approximated by the distribution $\text{Pois}(np)$. Let's apply this approximation to a real-world problem!

In the popular role-playing game *Dungeons & Dragons*, a 20-sided dice is commonly rolled. When a '1' is rolled on the dice, that's called *critical failure*, which usually results in something terrible happening. Suppose that in the course of a game, a 20-sided dice is rolled $n = 100$ times. Let X be the number of critical failures during the game. The probability of a critical failure is $p = 1/20 = .05$, so

$$X \sim \text{Bin}(100, .05).$$

(i) Suppose we wish to approximate X with a $Y \sim \text{Pois}(\lambda)$. What would be a good value of λ?

(ii) Using the `R` functions `rpois()` and `rbinom()`, generate 100,000 samples from the distributions of X and Y, which you should save as vectors `X_samples` and `Y_samples` respectively. Then use the following code to plot two overlapping histograms:

```
hist(X_samples, probability=TRUE, breaks=20, col=rgb(1,0,0,0.4), border="red",
     main="Critical Failures: Binomial (X) vs Poisson (Y)",
     xlab="Number of critical failures in 100 rolls", xlim = c(0,20))

hist(Y_samples, probability=TRUE, breaks=20, col=rgb(0,0,1,0.4), border="blue",
     add=TRUE)

legend("topright", legend = c("approx. distribution of X",
       "approx. distribution of Y"), fill=c(rgb(1,0,0,0.4), rgb(0,0,1,0.4)))
```

(iii) Using your Poisson approximation, what is the approximate probability of 2 or fewer critical failures during the course of the game?

Problem 10. An urn contains 5 balls, three red balls and two blue balls:

We consider the problem of sampling 3 balls from the urn, drawing the balls "without replacement". This means we draw one ball at random, then draw another ball at random, and then draw a third ball at random, without ever putting any of the balls back into the urn.

For $k = 1, 2$, and 3, we will use the notation R_k to denote the event that the k^{th} drawn ball is red, and B_k to denote the event that the k^{th} drawn ball is blue. Obviously, $\mathbb{P}[R_1] = 3/5$ and $\mathbb{P}[B_1] = 2/5$.

(a) Compute the conditional probabililities $\mathbb{P}[R_2 \mid B_1]$ and $\mathbb{P}[R_2 \mid R_1]$.

(b) Use the **Law of Total Probability** (Theorem 2.8 in the textbook, p70) and your answer to part (a) to compute $\mathbb{P}[R_2]$.

(c) If E and F are events, we use the notation EF to denote the event that both E and F occur (i.e., $EF = E \cap F$). Compute the probabilities of the four events R_1R_2, R_1B_2, B_1R_2 and B_1B_2.

(d) Use the Law of Total Probability and your answer to part (c) to compute $\mathbb{P}[R_3]$.

(e) In the remainder of this problem, we will compute the expected proportion of red balls among our 3 draws. To do this, we will introduce a standard technique: the use of indicator functions.

Given an event E, the ***indicator function of E*** is the function

$$\mathbf{1}_E = \begin{cases} 1 & : \text{ the event } E \text{ occurs} \\ 0 & : \text{ the event } E \text{ does not occur.} \end{cases}$$

Indicator functions are random variables. Taking $E = R_k$, we have

$$\mathbf{1}_{R_k} = \begin{cases} 1 & : \text{ the } k^{\text{th}} \text{ ball drawn is red} \\ 0 & : \text{ the } k^{\text{th}} \text{ ball drawn is blue.} \end{cases}$$

A random variable is ***discrete*** if it can assume only a finite or countably infinite number of distinct values.

If X is a discrete random variable, and $S_X \subseteq \mathbb{R}$ is the set of possible values that X can take, then the ***expectation*** of X, denoted $\mathbb{E}[X]$, is defined as

$$\mathbb{E}[X] := \sum_{x \in S_X} x\mathbb{P}[X = x],$$

provided that this sum converges absolutely.

Using the above definition of expectation, prove that $\mathbb{E}[\mathbf{1}_{R_k}] = \mathbb{P}[R_k]$ for $k = 1, 2, 3$.

(f) Observe that

$$(\#\text{ of red balls in 3 draws}) = \mathbf{1}_{R_1} + \mathbf{1}_{R_2} + \mathbf{1}_{R_3},$$

and hence

$$(\text{the proportion of red balls in 3 draws}) = \frac{\mathbf{1}_{R_1} + \mathbf{1}_{R_2} + \mathbf{1}_{R_3}}{3} \tag{1}$$

The **linearity of expectation** says that if X, Y are random variables, and a, b are scalars, then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$. Use the linearity of expectation and Eq. (1) to compute the expected proportion of red balls in 3 draws. *(Note: if you've done all parts of this problem correct, you'll get 3/5.)*