Math 472: Homework 01

Due Wednesday, September 10 (at the beginning of class)

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/`

(c) Do the `R` tutorial

(d) DO `https://swcarpentry.github.io/shell-novice/` or `https://carpentries-incubator.github.io/git-novice-branch-pr/` or `https://cecileane.github.io/computingtools/pages/notes0922-markdown.html`

(e) Write an R function to generate and plot a 1 dimensional brownian motion

Remark 2. The next problem is intended to introduce/review some central concepts in probability. For the remainder of this course, any terms defined using *this format* are precise definitions and should be memorized (they are fair game for in-class quizzes) Any terms that **are bolded** are important and should be reviewed if you are not familiar with them.

Problem 3. 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 expectations 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.)*

Problem 4 (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 5 (Exercise 1.9 in the textbook)**.** Resting breathing rates for college-age students are approximately normally distributed with mean 12 and standard deviation 2.3 breaths per minute. What fraction of all college-age students have breathing rates in the following intervals?

(a) 9.7 to 14.3 breaths per minute

(b) 7.4 to 16.6 breaths per minute

(c) 9.7 to 16.6 breaths per minute

(d) Less than 5.1 or more than 18.9 breaths per minute

Problem 6 (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 7. The *power set* of $\mathbb{R}$, denoted $2^{\mathbb{R}}$ is the set

$$2^{\mathbb{R}} := \{A : A \subseteq \mathbb{R}\}.$$

Let $P : 2^{\mathbb{R}} \to \{0, 1\}$ be the function defined by

$$P(A) := \begin{cases} 1 & : \quad A \text{ is uncountable} \\ 0 & : \quad \text{otherwise} \end{cases}$$

Verify that P is a probability measure on $2^{\mathbb{R}}$.

Problem 8. A discrete random variable X is said to be *binomially distributed* with parameters n (*number of trials*) and p (the *success probability*) if it has probability mass fucntion

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k \in \{0, 1, 2, \ldots, n\}$. Fix $\lambda > 0$. Suppose that $p = \lambda/n$. Show that

$$\lim_{n\to\infty} f_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

This shows that, in some sense, the Poisson distribution is the limit of the binomial distribution.

Problem 9. Prove that for any $x \in \mathbb{R}$,

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

Problem 10. Prove that if $P(\cdot)$ is a probability function and that B is an event with $P(B) > 0$, then the function $P(\cdot \mid B)$ is also a probability function.

Problem 11. 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 12. An example of a random variable whose expected value does not exist is the *Cauchy random variable*, that is, one with pdf

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \; x \in \mathbb{R}.$$

(a) Show that f_X is a probability density function.

(b) Show that $E|X| = +\infty$.