18.05 Problem Set 3, Spring 2022

Problem 1. (25: 10,5,10 pts.) **Independence**
Three events A, B, and C are *pairwise independent* if each pair is independent. They are *mutually independent* if they are pairwise independent and if, in addition,

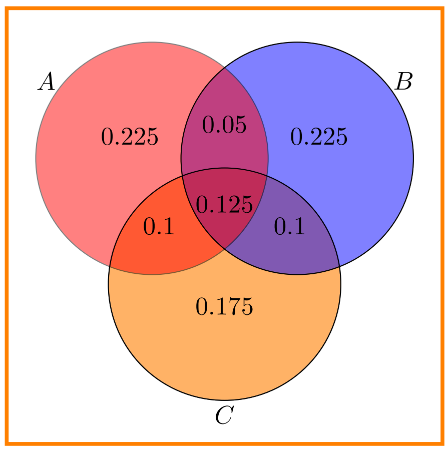
$$P(A \cap B \cap C) = P(A)P(B)P(C). \qquad (1)$$

(a) Suppose we roll two 6-sided die. Consider the events:

$A =$ 'odd on die 1' $\qquad B =$ 'odd on die 2' $\qquad C =$ 'odd sum'

Are A, B, and C pairwise independent? Are they mutually independent?

(b) Consider the Venn diagram below. A, B and C are the overlapping circles and the probabilities of each region are as marked. Does equation (1) hold. Are the events A, B, C mutually independent?



(c) Consider a litter of n puppies. What value(s) of n makes the events 'the litter has puppies of both sexes' and 'there is at most one female' independent.

Problem 2. (25: 5,5,10,5 pts.) **What does the data say?**
Suppose there is an experimental medical treatment for a cancer that, if untreated, is nearly always fatal within 12-15 months. The doctors enroll 5000 patients in a study in which each patient is given the treatment and followed for 5 years. Let X be the length of time a random patient given the treatment survives. (If a patient is still alive at the end of the study, then $X = 5$ for this patient.)

As the statistician it is your job to analyze the data.

To load the data into R you should do the following:

1. Download the data `mit18_05_s22_ps3prob2-data.r`. You can find this on our course website on the page with R code.
2. Put this file in your 18.05 R directory.
3. In R studio make sure the working directory is set to your 18.05 R directory.
4. Run the commands:

```
> source('mit18_05_s22_ps3prob2-data.r')
```

```
> x = get_prob2_data()
```

The variable `x` should now hold an array of the 5000 data points.

(a) Use R to compute the mean, variance and standard deviation of the data.

(b) Use the `hist` command to get R to plot a frequency histogram of the data. Set the histogram so each bin has width 0.1 years. Print the histogram and turn it in with the pset. The `hist()` command was introduced in Studio 3. There is also a short tutorial on using R to plot histograms on our class R page.

(c) Using your answers in (a) and (b), write a short paragraph summarizing the data in a useful way.

(d) Based on the (c), what are your conclusions about the effectiveness of the treatment? What recommendations would you make for avenues of further research?

Problem 3. (30: 10,10,10 pts.) **Dice**
Let X be the result of rolling a fair 4-sided die. Let Y be the result of rolling a fair 6-sided die. Let Z be the average of X and Y.

(a) Find the standard deviation of X, of Y, and of Z.

(b) Carefully graph the pmf and cdf of Z.

(c) Here is a gambling game: You win $2X$ dollars if $X > Y$ and lose 1 dollar otherwise. After playing this game 60 times, what is your expected total gain (positive) or loss (negative)?

Problem 4. (20: 5,5,5,5 pts.) **Two scoops**
Boxes of Raisin Bran cereal are 30cm tall. Due to settling, boxes have a higher density of raisins at the bottom ($h = 0$) than at the top ($h = 30$). Suppose the density (in raisins per cm of height) is given by $f(h) = 40 - h$.

(a) How many raisins are in a box?

(b) Let H be the height of a random raisin. Find and graph the pdf $g(h)$ of H.

(c) Find and graph the cdf $G(h)$ of H.

(d) What is the probability that a random raisin is in the bottom third of the box?

Problem 5. (20: 5,5,10 pts.) **Gallery of continuous random variables.**
The `pnorm()` function on R gives the cdf of the normal distribution, e,g, if $X \sim \mathrm{N}(\mu, \sigma^2)$ then $\texttt{pnorm}(x, \mu, \sigma) = P(X \leq x) = F_X(x)$.

(a) Suppose Z is a standard normal random variable. Use R to compute

(i) $P(Z \leq 0)$, (ii) $P(Z > 1.5)$ (iii) $P(|Z| < 1.5)$.

(b) Let $X \sim \mathrm{N}(\mu, \sigma^2)$ where $\mu = 2$ and $\sigma = 3$. Use R to compute

(i) $P(X \leq \mu)$, (ii) $P(X - \mu > 1.5\sigma)$ (iii) $P(|X - \mu| < 1.5\sigma)$.

(c) Let $Y \sim \exp(\lambda)$. Compute the cdf of Y by integrating the pdf. What is the probability $Y \leq 1/\lambda$? (You need to do an integration, but you can check your work numerically using the `pexp()` function in R.)

Problem 6. (20: 5,5,5,5 pts.) **Birth day**
The length of human gestation is well-approximated by a normal distribution with mean $\mu = 280$ days and standard deviation $\sigma = 8.5$ days.

(a) Graph the corresponding pdf and cdf. You should do this using the dnorm, pnorm and plot commands in R. Print the results and turn them in with the pset.

(b) Suppose your final exam is scheduled for May 18 and your pregnant professor has a due date of May 25. Find the probability she will give birth on or before the day of the final.

(c) Find the probability she will give birth in May sometime after the exam. (Assume this means from the start of May 19 to the end of May 31.)

(d) The professor decides to move up the exam date so there will be a 95% probability that she will give birth afterward. What date should she pick?