18.05 Problem Set 3, Spring 2018

(due outside 4-174 Monday, Feb. 26 at 9:30 AM)

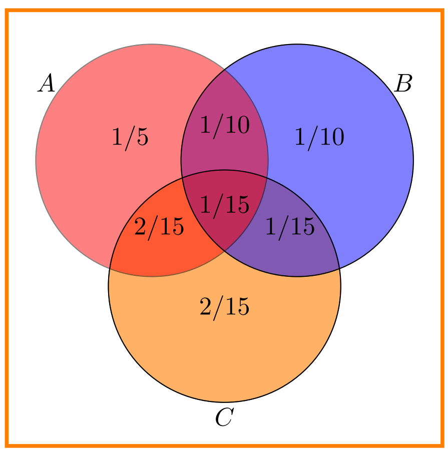
Problem 1. (10 pts.) **Independence.**

(a) Suppose we have three events A, B, C. Suppose also that A is independent of B, and B is independent of C. Does it follow that A is independent of C? Either prove this is the case or else give an example where it is not true.

(b) Three events A, B, and C are pairwise independent if each pair is independent. They are mutually independent if they are pairwise independent and in addition

$$P(A \cap B \cap C) = P(A)P(B)P(C). \tag{1}$$

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) We have a bent coin with probability of heads $= 1/3$. Suppose we toss the coin three times. Assume the tosses are independent and consider the events:

$A =$ 'the first toss is heads' $B =$ 'the second toss is heads' $C =$ 'the total number of tails is 2'.

Are the three events mutually independent?

(d) Suppose that events D and E are independent. Show that D and E^c are also independent.

Hint: We know $P(E \cap D) = P(E)P(D)$. We must show $P(E^c \cap D) = P(E^c)P(D)$. What's the union of the two events $E \cap D$ and $E^c \cap D$?

Problem 2. (10 pts.) **On a roll.** The following game is played with a four-sided die: a turn consists of a player rolling the die until either they roll a 1 or they have rolled 5 times without getting a 1.

(a) Let X be the number of rolls on a given turn (including the roll of 1 that ends the turn). Compute the mean, variance and standard deviation of X.

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

(c) A casino turns this into a gambling game: a player pays \$1 to take a turn; let X be their number of rolls. Then the house takes a turn; let H be the house's number of rolls. If $X > H$ then the house pays out \$2 to the player; otherwise they pay out nothing.

What is the player's expected gain (positive) or loss (negative)?

Problem 3. (10 pts.) **Analyzing data.**
If untreated a certain disease is nearly always fatal within 1–2 years of diagnosis. An experimental treatment was tested in a study of 5000 patients. In the study each patient was given the treatment and followed for 5 years. The data consists of the length of time each patient survived. (If a patient was still alive at the end of the study, then their survival length is recorded as 5 years.) Your job is to analyze the data from this experiment.

To load the data into R you should do the following:
1. Download the data `ps3prob3-data.r` from the class web site (link "Here is R").
2. Put this 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('ps3prob3-data.r')
> x = getprob3data()
```

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

(a) Use R to compute the mean and variance 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. There is a short tutorial on using R to plot histograms on our class web site (go to the same page as for ps3prob3-data.r and look in the tutorials section.)

(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 4. (10 pts.) **Winter in New England**. My house has a flat roof, so I decided to push the snow off of it. While up there I noticed that because of the way our neighbor's house blocks the wind the snow was much deeper on the west side of the roof than the east. I estimated that the depth of the snow was given by the formula $h(x) = 8 + x^2/40$ inches, where $0 \le x \le 40$ is the number of feet from the east edge of the roof.

(a) Assume I am equally likely to be at any spot on the roof and let Y be the depth of the snow at that spot. Find and graph both the pdf $f(y)$ and cdf $F(y)$ of Y.

Hint: If X is my distance from the east edge, then X follows a uniform random variable. Y is a transform of X.

(b) While I was up there I dropped my phone. What is the probability that it fell into snow less than 15 inches deep?

Problem 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 $\texttt{pnorm}(x, \mu, \sigma) = P(X \le x) = F_X(x)$.

(a) Suppose Z is a standard normal random variable. Use R to compute
(i) $P(Z \le 0)$ (ii) $P(Z > 1.5)$ (iii) $P(|Z| < 1.5)$.

(b) Let $X = \mathrm{N}(\mu, \sigma^2)$ where $\mu = 2$ and $\sigma = 3$. Use R to compute
(i) $P(X \le \mu)$ (ii) $P(X - \mu > 1.5\sigma)$ (iii) $P(|X - \mu| < 1.5\sigma)$.
Explain why you could have also computed this directly from your answers to part (a)

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

Problem 6. (10 pts.) **Galton-Binet-Stanford.**
IQ scores in the general population are well-approximated by a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$.

(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) What percentage of the population has an IQ lower than 92?

(c) What IQ cutoff should you use if you want to have a 95% probability that a randomly chosen person has IQ above the cutoff?