18.05 Problem Set 3, Spring 2014

Problem 1. (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 in addition

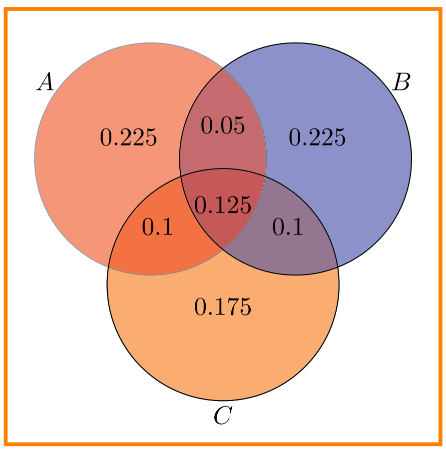
$$P(A \cap B \cap C) = P(A)P(B)P(C). \quad (1)$$

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

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

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) For families with n children, the events the family has children of both sexes and there is at most one girl are independent. What is n?

Problem 2. (10 pts.) **R simulation.** 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 put the data in a vector x you need to do the following.

First download

http://ocw.mit.edu/ans7870/18/18.05/s14/r-code/ps3prob2data.r

and put the file in your R working directory. Then give the following R commands.

```
> source('ps3prob2data.r')
> x = getprob2data()
```

(a) Compute the mean and standard deviation of the data.

(b) 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.

(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. (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) 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. (10 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. (10 pts.) **The new normal.** Recall that the normal distribution $N(\mu, \sigma^2)$ has pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The standard normal distribution $N(0, 1)$ has mean 0 (by symmetry), variance 1 (as we'll prove next week), and pdf $\phi(z)$ given by setting $\mu = 0$ and $\sigma = 1$ above. The cdf is denoted $\Phi(z)$ and does not have a nice formula. In this problem, we ll show that scaling and shifting a normal random variable gives a normal random variable. Suppose $Z \sim N(0, 1)$ and $X = aZ + b$.

(a) Compute the mean μ and variance σ^2 of X.

(b) Express the cdf $F_X(x)$ of X in terms of Φ and then use the chain rule to find the pdf $f_X(x)$ of X.

(c) Use (b) to show that X follows the $N(b, a^2)$ distribution.

(d) Use (a) and (c) to conclude that the $N(\mu, \sigma^2)$ distribution has mean μ and variance σ^2.

Problem 6. (10 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.

Suppose your final exam is scheduled for May 18 and your pregnant professor has a due date of May 25.

(b) 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.

(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?