Instructions: Submit one .pdf file that includes the answers to both R questions and conceptual questions. Also make sure you submit your .R file with the R code you used to get the answers. Use your R code from last week's homework as a template for the R script this week. Make sure to use comments in your code to highlight where a questions starts and ends.

Please feel free to discuss questions on the discussion board.

R Questions

1. (6 points) Simulation and estimation in R

   (a) The R function `runif()` generates random numbers from the continuous Uniform distribution. For example, `runif(1, min = 0, max = 1)` generates one random number from a Uniform(0,1) distribution, i.e. one random number between 0 and 1 with uniform probability.

      i. Create a histogram of 10000 draws from a Uniform(0, 1) distribution.

      ii. Estimate the probability a random variable with this distribution is between 0.5 and 0.75

   (b) The `rbinom()` function in R simulates random variables from a Binomial Distribution. Recall a Binomial random variable is the number of successes in n trials where the probability of success is π. In `rbinom()` n is specified using the `size` argument and π by the `prob` argument. Confusingly, `rbinom()` has an argument called `n`, but this how many times we would like to simulate from a Binomial(n, π) not the value of n.

      So, for example, `rbinom(n = 1, size = 10, prob = 0.5)` will simulate one random variable that has a Binomial(10, 0.5) distribution, whereas `rbinom(n = 10, size = 1, prob = 0.5)` will simulate 10 random variables each with a Binomial(1, 0.5) distribution.

      In the U.S. having a baby girl isn't as equally likely as having a baby boy, in fact if we think of "having a baby girl" as a success in a single "having a baby" trial, then the probability of success is about 0.49 (see CIA World FactBook).

      i. Estimate the probability that for a family who has two children both will be girls (Hint: Estimate $P(X = 2)$ when $X \sim Binomial(2, 0.49)$).

      ii. Estimate the probability that for a family who has four children all are girls.

      iii. Estimate the probability that for a family with two children one of whom is a boy the other child is a girl (Hint: This is tricky because it is a conditional probability. Start by just using realizations where one child is a boy (i.e. $X = 0$, or 1), what proportion have a boy and a girl (i.e. $X = 1$)?)

      iv. Why can't we estimate the probability that a randomly chosen family has at least one girl? What additional information would we need to estimate this?

Conceptual Questions

Answer any two of the following five short answer questions.

2. (2 points) Consider the rolling of a **fair**, six-sided die.

   (a) Why do we consider this to be random?

   (b) Identify the outcomes.

   (c) Identify two possible events.

(d) Identify the probability associated with each outcome. Why did you choose these particular probabilities?

(e) Define one random variable that may be associated with the rolling of the die.

(f) If we were able to roll the die infinitely many times, what would happen to the relative frequencies of each outcome? What is the name of this phenomenon?

3. (2 points) The integrity of twenty thin metal plates is tested by bending them until they break. The number of pounds per square centimeter (cm) of force required to break each plate is measured and recorded. Five of the plates break with 50 lbs/cm^2 of pressure or less, while the remaining fifteen plates break with more than 50 lbs/cm^2 of pressure.

   (a) What is the proportion of plates that break under the lower amount of pressure?

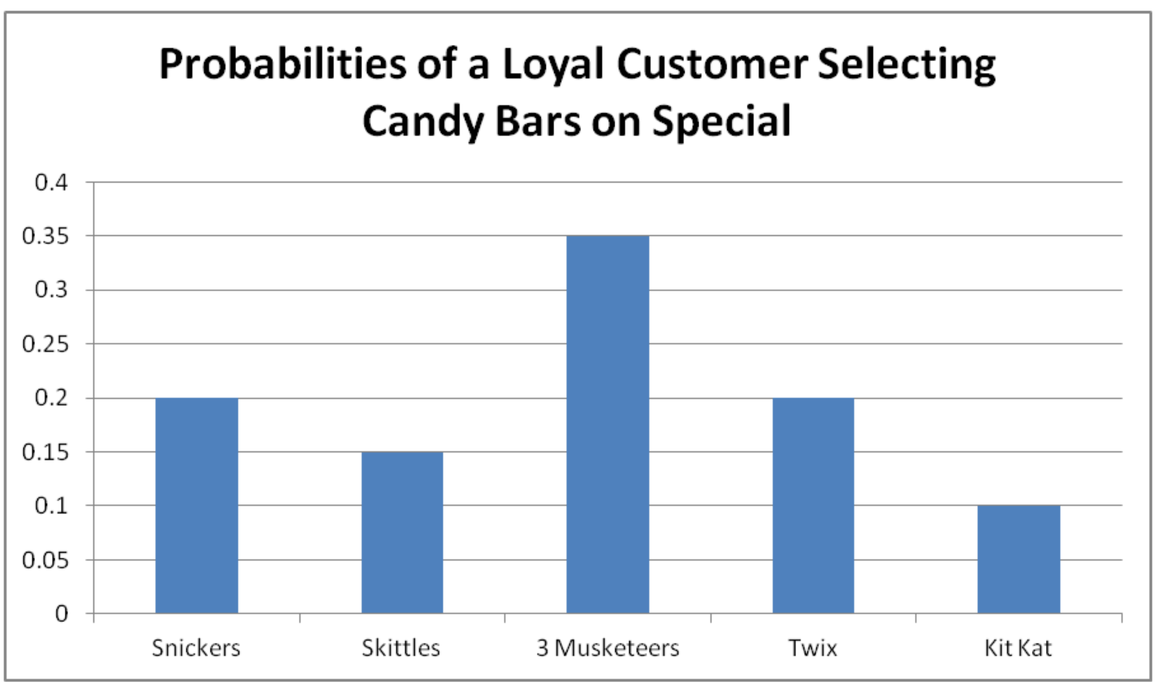
   (b) Is the proportion you calculated in step A equal to the probability of a plate breaking under 50 lbs/cm^2 of pressure or less? Why or why not? If not, what would you need to do to find the exact probability?

   (c) As more plates are tested, what will happen to the proportion of plates that break under 50 lbs/cm^2 of pressure or less? What is this phenomenon called?

   (d) Assume that the probability of a plate breaking under 70 lbs/cm^2 of pressure or less is 0.66, the probability of a plate breaking under between 50 and 80 lbs/cm^2 of pressure is 0.73, and the probability of a plate breaking under between 50 and 70 lbs/cm^2 of pressure is 0.5. What is the probability of a plate breaking under more than 80 lbs/cm^2 of pressure? (Hint: Draw a picture)

   (e) Now assume that one thousand plates are tested and 108 of them break under more than 80 lbs/cm^2 of pressure. Does this contradict your answer in part D? Why or why not?

4. (2 points) There are five candy choices for sale at a local candy shop: Twix, KitKat, Snickers, Skittles and 3 Musketeers. For a random customer, the probabilities of choosing each candy bar are represented by the bar chart shown here:



(a) What is the probability that the customer will buy a Twix or KitKat? What is the probability that the customer will buy Snickers, Skittles or 3 Musketeers? What is the probability that this customer will not buy any candy?

(b) Let X be a random variable representing the candy bought—you should assign the candy choice outcomes to the integers 1 through 5 in the same order as on the bar chart (left to right). Write the probability distribution of X in tabular format.

5. (2 points) George thinks it is his lucky day. He walks into a casino and decides to play a dice game. He must roll two **fair** six-sided dice at once. If he rolls two 2's, he wins one hundred dollars. However, it costs ten dollars to roll the dice one time. George has rolled the dice nineteen times and has not yet rolled two 2's. He is behind in money, but feels that since he has not rolled two 2's yet, it is bound to happen soon so he can make at least *some* of his money back. George asks you what he should do. Statistically speaking, what should you tell him and why? Mention the Law of Large Numbers in your answer.

6. (2 points) A baseball player is at batting practice. He decides to swing the bat twenty five times. On any given swing, the probability that he hits the ball is 0.3. Explain how the four properties of a binomial variable hold in this example. Assume that the player's current swing is not affected by any of his previous swings.