R computing for Business Data Analytics

Homework 2

Due date: October 27, 2014

Please e-mail your homework (.pdf) and the associated *R* code (.*R*) to hchuang.om@gmail.com.

The email title must be R_HW2_GroupName. NO late homework will be accepted.

Q1. (20%) Two six-sided dice are thrown sequentially, and the face values that come up are recorded.

(a) List the sample space.

(b) List the elements that make up the following events: (1) *A*=the sum of the two values is at least 5 (2) *B*=the value of the first die is higher than the value of the second, (3) *C*=the first value is 4.

(c) List the elements of the following events: (1) $A \cap C$, (2) $B \cup C$, (3) $A \cap (B \cup C)$.

(d) Based on the classical approach, derive $P(A \cap C)$, $P(B \cup C)$, and $P(A \cap (B \cup C))$.

(e) Use the *sample* function to simulate the throwing of two dice 1,000 times. Compute $P(A \cap C)$, $P(B \cup C)$, and $P(A \cap (B \cup C))$ from the 1,000 runs (i.e., the relative frequency approach). How different are the results from (d) (show me the *R* code in the .*R* file)?

Q2. (20%) Use the *choose* function in *R* to answer (c), (d), and (e)

(a) Prove Bonferroni's inequality: $P(A \cap B) \geq P(A) + P(B) - 1$

(b) The first three digits of a university telephone exchange are 452. What is the probability that a randomly selected university phone number contains seven distinct digits?

(c) How many different meals can be made from four kinds of meat, six vegetables, and three starches if a meal consists of one selection from each group?

(d) A committee consists of five Chicanos, two Asians, three African Americans, and two Caucasians. A subcommittee of five is chosen at random. What is the probability that all the ethnic groups are represented on the subcommittee?

(e) A drawer of socks contains seven black socks, eight blue socks, and nine green socks. Two socks are chosen in the dark. What is the probability that they match? What is the probability that a black pair is chosen?

Q3. (10%) Simulating Blackjack (classical versus relative)

(a) In the poker game Blackjack, let each of the 4 aces denote 11 points and each of the 16 cards > 10 (i.e., 10, J, Q, K) denote 10 points. Suppose someone picks 2 cards randomly out of a deck of cards (52 cards total), what is the probability of getting 21 points (i.e., Blackjack)?
Use the *choose* function to obtain the answer.

(b) Now, let the numbers 1-52 represent a deck of cards. Assign the numbers 1-4 to the four aces and the numbers 37-52 to the 16 cards > 10.
Use the *sample* function and the *for* loop to simulate the random draw of 2 cards **50,000** times (hint: with or without replacement?). Create a variable *success* that represents the number of times you hit 21 points (i.e., Blackjack). Divide success by 50,000 to obtain the relative probability. Is the relative probability close to the classical one from part (a)?
What will happen to the relative probability if you only simulate the game for 50 times?

Q4. (20%) Conditional probability, law of total probability, and Bayes theorem.

(a) Two dice are rolled, and the sum of the face value is six. What is the probability that at least one of the dice camp up a three?

(b) A fire insurance company has 10% high-risk, 20% medium-risk, and 70% low-risk clients, with respective probabilities 0.02, 0.01, and 0.0025 of filing claims within a given year. What proportion of the claims filed each year come from high-risk clients?

(c) A company that manufactures digital camcorders produces a basic model and a deluxe model. Over the past year, 40% of the camcorders sold have been of the basic model. Of those buying the basic model, 30% purchase an extended warranty, whereas 50% of all deluxe purchasers do so. If you learn that a randomly selected purchaser has an extended warranty, how likely is it that he/she has a basic model?

(d) A factory runs three shifts. Each shift produces 100 items per day. In a given day, shift one has 1% defective items, shift two has 2% defects, and shift three has 5% defects. What is the percentage of defective items? If an item is defective, what is the probability that it was produced by shift three?

Q5. (15%) A bank has five desks 1, 2,…,5. The percentages of customers visiting each desk are 20%, 30%, 10%, 15%, and 25%. Among customers for each desk, the percentages that customers would stay longer than 30 minutes are 40%, 60%, 20%, 80%, and 90%.

(a) Draw a tree diagram to model the process.

(b) What is the overall probability of customers staying longer than 30 minutes?

(c) If a newly entering customer stays longer than 30 minutes, what are the posterior probabilities that she/he visits desks 1, 2,…,5? (i.e., find $P(D_1|E)$, $P(D_2|E)$,…, $P(D_5|E)$ and make sure that they sum to 1)

Q6. (15%) Assume that grades are posted according to the last 4 digits of one's ID. Use *R* to

(a) Find that the actual probability that at least two students in a class of 100 share the same ID.

(b) Simulate the last 4 digits for 100 students 5,000 times. How many times do you find at least two students have the same ID? Divide the number by 5,000, what is the fraction?

(c) What is the smallest class enrollment (i.e., number of students) for which the probability that at least two students have the same ID numbers is at least 0.5?