MPCS 53110 Foundations of Computational Data Analysis — Winter 2017

Homework 2. This problem set is due Tuesday January 17 at 5:30 pm.

Reading: Probability & Statistics with R (= P&S): chapter 3, sections 3.3–3.4.

Written assignment:

- Solve the following "DO" exercises and homework problems "HW".
- "DO" exercises are strongly recommended to check your understanding of the concepts. **Do not submit them**.
- **Problems labeled "HW" must be submitted**.
- You are responsible for the material covered in **both** "DO" exercises and HW problems.

"Do" Exercises (*not* to be submitted):

1. **"DO"** Exercises 3.4.4, 3.4.5, and 3.4.6, P&S pages 144–145.

2. **"DO"** At least one-half of an airplane's engines are required to function in order for it to operate. If each engine independently functions with probability p, for what values of p is a 4-engine plane more likely to operate than a 2-engine plane?

3. **"DO"** If X is a binomial random variable with parameters n and p, where $0 < p < 1$, show that
   - $P(X = k + 1) = p/(1 - p) \times (n - k)/(k + 1) \times P(X = k)$, $k = 0, 1, \ldots, n - 1$.
   - As k goes from 0 to n, $P(X = k)$ first increases and then decreases, reaching its largest value when k is the largest integer less than or equal to $(n + 1)p$.

4. **"DO"** Approximately 80,000 marriages took place in the state of New York last year. Estimate the probability that for at least one of these couples:
   - both partners were born on April 30
   - both partners celebrated their birthday on the same day of the year

   State your assumptions.

5. **"DO"** If X is a Poisson random variable with mean λ, show that $P(X = i)$ first increases and then decreases as i increases, reaching its maximum value when i is the largest integer less than or equal to λ.

Homework Problems (to be submitted Tuesday January 17 at 5:30 pm):

- **Collaboration policy: If you work with others, indicate their names as part of your submission. You must answer each question by yourself without assistance. It is a violation of this policy to submit a solution that you cannot explain orally to the instructor/TAs.**
- **Looking for solutions to problems on the internet, even when acknowledged, is STRONGLY DISCOURAGED.**
- **Write out your work for every "theory" problem. If you just write your answer without showing your work, you will not receive credit.**

1. **HW** Let X be a random variable with PDF $f_X(x) = 0.1 \exp(-0.1x)$ for $x > 0$ and 0 otherwise.
   - Find $E[X]$. Show your work.
   - Find $\mathrm{Var}(X)$. Show your work.

   (1 point each)

2. **HW** Let a and b be positive integers with $a \le b$, and let X be a random variable that takes as values, with equal probability, the powers of 2 in the interval $[2^a, 2^b]$.
   - Find $E[X]$. Show your work.
   - Find $\mathrm{Var}(X)$. Show your work.

   (2 points each)

3. **HW** Alice throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of Alice's hit from the center.
   - Find the PDF, the expected value, and the variance of X. Show your work. (3 points)
   - The target has an inner circle of radius t. If $X \le t$, Alice gets a score of $S = 1/X$. Otherwise her score is $S = 0$. Find the CDF of S. Show your work. Is S a continuous random variable? Explain your answer. (3 points)

4. **HW** We are often interested in how good of an approximation one distribution is to another. One way to measure the quality of an approximation to a discrete probability distribution is to treat both probability mass functions as vectors, and measure their Euclidean distance. The function linked here dist.r does this.
   - Write a command to give the probability mass function of the binomial random variable with $n = 8$ and $p = 13/34$ as a vector. (We just want the probabilities in order, not the x-values that they correspond to.) Write a function that, given a positive integer input x, gives the probability mass function of the hypergeometric random variable with m = 13*x, n = 21*x, and k = 8, as a vector. (You may find dbinom() and dhyper() convenient for this.) (2 points)
   - Plot the distance between binom(n=8,p=13/34) and hyper(m=13*x,n=21*x,k=8) for x = 1 to 1000. (2 points)
   - Find a simple but accurate model for the distance between binom(n=8,p=13/34) and hyper(m=13*x,n=21*x,k=8) that does not require you to explicitly compute any pmf's. Implement your model as an R function that takes as input a positive integer x, and outputs an estimate for the distance between binom(n=8,p=13/34) and hyper(m=13*x,n=21*x,k=8). Hint: It may help to use the lm() function to find linear regressions between different transformations of the variables. Try taking logarithms of one or both variables, or raising one or both variables to different powers. Note that if you want to include mathematical operations directly in an R formula, you will need to surround the expression with the I() function. (6 points)

5. **HW** In this problem you will construct and analyze a basic model of a queue, processing inputs that arrive at random times. The queue operates in two alternating phases. In Phase I, inputs arrive in the queue. The number of inputs attempting to enter the queue is a Poisson random variable with mean λ. However, the queue has a maximum size b, so if more inputs arrive than can fit into the queue, the queue overflows and the inputs that cannot fit into the queue are discarded. In Phase II, some or all of the inputs at the head of the queue are removed and processed. The maximum number of inputs that can be processed at once is c, an integer with $0 < c \le b$. If there are fewer than c inputs in the queue, they are all processed; otherwise, exactly c inputs are processed.
   - Write an R function that takes an input λ, b, and c, simulates this queue model, and returns the numer of iterations of Phase I before the first overflow. (6 points)
     Hint: see ?rpois. Attached are some tests on the model answer: queue model tests; you should get similar results.
   - For the special case that $b = c$, give an expression in terms of λ and b for the probability of an overflow. (4 points)
   - If $b = c$, what is the expected output of your function in terms of λ and b? Explain your answer in terms of random variables we have studied. (2 points)

Gerry Brady, David Ramsey, Liwen Zhang, and Colin Hudler
Wednesday January 11 14:16:34 CST 2017