MPCS 53110 Foundations of Computational Data Analysis — Winter 2017

Homework 3. This problem set is due Tuesday January 24 at 6:00 pm.

Reading: Probability & Statistics with R (= P&S): chapter 3, section 3.5; chapter 4, sections 4.1–4.3.

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"** Section 3.2: exercise 10, page 110.

2. **"DO"** Section 3.3: exercise 7, page 122.

3. **"DO"** Section 3.4: exercises 2, 12, 18, 19, and 21, pages 144–147.

4. **"DO"** Section 3.5: exercises 1, 2, 5, and 7, pages 156–157.

5. **"DO"** Section 4.2: exercises 1, 3, and 4, pages 165–166.

Homework Problems (to be submitted Tuesday January 24 at 6:00 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** We say that x_α is the $100(1 - \alpha)$ percentile of the CDF F_X if
   $$F_X(x_\alpha) = 1 - \alpha.$$
   Find x_α for the CDF having PDF
   $$f_X(x) = 2\exp(-2x) \text{ for } x \geq 0 \text{ and } 0 \text{ otherwise}.$$
   Show all your work and explain your answer. (5 points)

2. **HW** Evidence concerning the guilt or innocence of a defendant in a criminal investigation can be summarized by the value of an exponential random variable X whose mean μ depends on whether the defendant is guilty. If innocent, $\mu = 1$; if guilty, $\mu = 2$. The deciding judge will rule the defendent guilty if $X > c$ for some suitably chosen value of c.
   - If the judge wants to be 95% certain that an innocent man will not be convicted, what should be the value of c? Show your work and explain your answer. (3 points)
   - Using the value of c found above, what is the probability that a guilty defendant will be convicted? Show your work and explain your answer. (2 points)

3. **HW** The mean and standard deviation on your algorithms exam were 60 and 20, respectively; the mean and the standard deviation on your data analysis exam were 55 and 10, respectively. You scored 70 on the algorithms exam and 62 on the data analysis exam. Assume that the two histograms of test scores are approximately normal.
   - On what exam is your percentile score the highest? Show your work and explain your answer. (2 points)
   - What percentage of the scores on the algorithms exam are below your score? Show your work and explain your answer. (2 points)
   - What percentage of the scores on the data analysis exam are below your score? Show your work and explain your answer. (2 points)

4. **HW** A product is classified according to the number of defects it contains and the factory that produces it. Let X_1 and X_2 be the random variables that represent the number of defects per unit (taking on possible values of 0, 1, 2, or 3) and the factory number (taking on possible values of 1 or 2), respectively. The entries in the following table represent the joint PMF of a randomly chosen product.

$X_1 \backslash X_2$	1	2
0	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{3}{16}$	$\frac{1}{8}$
3	$\frac{1}{8}$	$\frac{1}{4}$

   - Find the marginal PMF's of X_1 and X_2. Show your work. (2 points)
   - Find $E[X_1]$ and $E[X_2]$; $\mathrm{Var}(X_1)$ and $\mathrm{Var}(X_2)$; and $\mathrm{Cov}(X_1, X_2)$. Show your work and explain your answers. (5 points)

5. **HW** As we saw in the intro to R, it is possible to simulate a large number of coin flips using the sample() function:

```
> sample(c("tails", "heads"), 20, replace = T)
[1] "tails" "tails" "tails" "tails" "tails" "heads" "tails" "heads" "heads"
[10] "tails" "tails" "tails" "tails" "heads" "tails" "tails" "heads" "heads"
[19] "tails" "heads"
```

If we represent "tails" as 0 and "heads" as 1, we can quickly find the number of heads in such a simulation:

```
> sum(sample(0:1, 20, replace = T))
[1] 13
```

Since each element of the sum is equal to 1 independently with probability 1/2, we see that this expression simulates a binomial random variable with $n = 20$ and $p = 1/2$. If we plotted a histogram of a sample of 1000 numbers generated from this expression, and a histogram of the expression rbinom(1000, n = 20, p = 1/2), we would get similar results.

For each of the distributions below, write a function in R that takes the specified parameters and outputs a random number with the corresponding distribution. The only source of randomness you may use is the sample() function, so e.g. you may not use the random number generators rbinom(), rgeom(), etc. You may also not use the PMF functions (dbinom(), dgeom(), etc.). You are encouraged to plot histograms of your function against histograms of the random number generators to test that your functions are behaving correctly.

- (a) Binomial distribution, with parameters n and p
- (b) Geometric distribution, with parameter p
- (c) Hypergeometric distribution, with parameters M_1, M_2, and N

(12 points)

Gerry Brady, David Ramsey, Liwen Zhang, and Colin Hudler
Wednesday January 18 13:38:19 CST 2017