MPCS 53110 Foundations of Computational Data Analysis — Winter 2017

Homework 4. This problem set is due Tuesday January 31 at 5:30 pm.

Reading: Probability & Statistics with R (= P&S): chapter 4, sections 4.3–4.6.

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"** Show the following relationships:
    1. $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$
    2. $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)$

2. **"DO"** Section 4.3: exercise 3, page 178.

3. **"DO"** Section 4.4: exercise 8, page 191.

4. **"DO"** Section 4.5: exercise 6, page 198.

Homework Problems (to be submitted Tuesday January 31 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 and Y have the joint PDF:

    $$f(x, y) = 24\,x, \quad \text{if } 0 \le y \le 1 - 2x \text{ and } 0 \le x \le 1/2$$
    $$= 0 \text{ otherwise.}$$

    - Find the marginals f_X and f_Y by hand. Show your work. (3 points each)
    - Are X and Y independent? Why or why not? (1 point)
    - Find $E[X]$ and $E[Y]$ by hand. Show your work. (3 points each)
    - Find the conditional PDF of Y given X: $f_{Y\,|\,X\,=\,x}(y)$. Your answer should be a function of x. (2 points)
    - Find the regression function $E[Y \,|\, X = x]$. Show your work. Your answer should be a function of x. (3 points)
    - Plot the regression function and give the numerical value of $E[Y \,|\, X = 0.3]$. (2 points)
    - Use the law of total expectation (eq. 4.3.15, p. 173) to find $E[Y]$ by hand. Show your work. (3 points)

2. **HW** Let X and Y have the joint PDF:

$f(x, y) = (6/7)\,(x + y)^2 \quad 0 \le x \le 1, \quad 0 \le y \le 1;$
$\quad\quad = 0$ otherwise.

- By integrating over the appropriate regions, find
  - $P(X > Y)$
  - $P(X + Y \le 1)$
  - $P(X \le 1/2)$

  (4 points each)
- Find the marginal PDF's f_X and f_Y by hand. Show your work. (3 points each)
- Are X and Y independent? Why or why not? (1 point)
- Find the two conditional PDF's of X given Y and Y given X: $f_{X \mid Y=y}(x)$ and $f_{Y \mid X=x}(y)$ (2 points each)
- Find the covariance of X and Y, $\mathrm{Cov}(X, Y)$, by hand. Show your work. (4 points)
- Find the correlation of X and Y, $\mathrm{Corr}(X, Y) = \varrho_{XY}$, by hand. Show your work. (2 points)
- Find $E[Y \mid X = x]$ for $0 \le x \le 1$ by hand. Show your work. (3 points)
- Comment on the appropriateness of the linear (Pearson's) correlation coefficient ϱ_{XY} as a measure of dependence between X and Y. (1 point)

3. **HW** Let X be the continuous uniform distribution on $[-1, 1]$ and let $Y = X^2$. Show that $\mathrm{Cov}(X, Y) = 0$ but that X and Y are not independent. (3 points)

4. **HW** In Example 5.3-4 in Akritas, p.219, a scheme for representing joint discrete random variables in R is described. Specifically, X and Y are sample spaces, and S is a data frame which is the result of calling expand.grid(X,Y), so all possible pairs of outcomes from X and Y are listed as rows of S. P is a data frame with a column called "xy", whose entries are the probabilities of each of the corresponding pairs of outcomes from S. (You should not expect that P will contain columns "x" and "y".)
   Write R functions that, when given the inputs S and P that describe a random variable according to this scheme, find each of the following:
   - The marginal probability distributions of X1 and X2.
   - The expected value of X1 and X2.
   - The variance of X1 and X2.
   - The covariance of X1 and X2.

   (4 points for each part)

Gerry Brady, David Ramsey, Liwen Zhang, and Colin Hudler
Wednesday January 18 23:11:21 CST 2017