Stat450: Take-home midterm exam

Fall 2016

Due Wednesday, November 2

This is a take-home exam. All work must be done individually. All work must be fully shown and justified to receive full credit: i.e., if you are asked to evaluate an integral, you must show work for this integral or appeal to well-known integrals to receive full credit. Appealing to calculator answers will not receive full credit.

Let X and Y be jointly continuous random variables with the following pdf:

$$f_{X,Y}(x,y) = \frac{x^{10}y^9}{4 \cdot 9!}e^{-xy},\ 1 \leq x \leq 5,\ y > 0.$$

1. The first part of this take-home requires evaluating the following quantities. If possible, you may appeal to any known facts about well-known distributions in any of your answers.

a) (3pts) Find and fully describe the marginal distribution of X.

b) (3pts) Find and fully describe the conditional distribution of $Y|X$.

c) (2pts) Find $E(Y|X)$.

d) (2pts) Find $Var(Y|X)$.

e) (2pts) Find $E(Y)$.

f) (2pts) Find $Var(Y)$.

g) (2pts) Find $Cov(X,Y)$.

h) (2pts) Find the theoretical R^2, which is the proportion of the total variability in Y that can be explained by X. The theoretical R^2 is defined to be:

$$R^2 = 1 - \frac{E_X(Var(Y|X))}{Var(Y)}.$$

2. The second part of the exam involves verifying responses from Part 1 with simulation studies. The entirety of Part 2 can be answered with R code. Write R code to obtain 1000 realizations of (X,Y) pairs, and calculate the following quantities based on these realizations. Verify that they are similar to the theoretical quantites you derived in Part 1. Include all R code along with relevant R output in your submission.

a) (2pts) Verify $E(Y)$, $Var(Y)$, and $Cov(X,Y)$.

b) (3pts) Make a scatterplot of Y versus X. Add lines for $E(Y|X)$ as well as $E(Y|X) \pm SD(Y|X)$ (see course notes for example).

c) (3pts) Does it appear from your scatterplot that X explains a significant amount of the variability in Y? Calculate the *empirical* (i.e. observed) R^2 by obtaining $\frac{1}{1000}\sum_{i=1}^{1000}(Y_i - E(Y_i|X_i))^2$. Use this along with the empirical $Var(Y)$ to find the empirical R^2. Verify that it approximates your answer to 1h.