"In the Name of God"

Engineering probabilities and Statistics **CE 40181**
Sharif University of Technology

Homework 4
Wednesday, Dec 13 2017

Reminders:

- **Typing the solutions has extra marks.**
- **Collaboration is permitted, but you must write solutions *by yourself without* assistance.**
- **Getting solutions from outside sources such as the Web or students not enrolled in the class is strictly forbidden.**
- **Late submissions will be treated according to the course policy.**

Problems:

1. The conditional covariance of X and Y , given Z, is defined by:

$$Cov(X, Y|Z) = E[(X - E[X|Z])(Y - E[Y|Z])|Z]$$

   (a) Show that:
$$Cov(X, Y|Z) = E[XY|Z] - E[X|Z]E[Y|Z]$$

   (b) Prove the conditional covariance formula:

$$Cov(X, Y) = E[Cov(X, Y|Z)] + Cov(E[X|Z], E[Y|Z])$$

   (c) Prove the conditional variance formula:

$$Var(X) = Var(E[X|Y]) + E[Var(X|Y)]$$

2. For jointly defined random variables X, Y and Z prove the followings:

   (a) $E[Y|Z] = E[E[Y|X, Z]|Z]$

   (b) $E[XY] = E[XE[Y|X]]$

   With the assumption of $E[Y|X] = 1$ prove that:

   (c) $Var(XY) \geq Var(X)$

3. Suppose that we have m computer engineers in our startup. When we have some tasks, they all start working on them at the same time. Each engineer chooses one task independently from the others to work on. If the probability of finishing the task for each engineer is p, and the number of tasks is n $\sim$ Poisson(λ), find average number of done tasks.

4. We are going to build a **tree**. At first we put vertex 1 as the root of the tree. Then in the i'th step ($2 \leq i \leq n$) we connect vertex i randomly to one of the previous vertices $(1, 2, ..., i-1)$ with the same probability. What is the expected value of the sum of the number of vertices in the path from vertex 1 to vertex n? (vertex 1 and n are also in that path)

   Note: A **tree** is an undirected graph in which any two vertices are connected by exactly one path. In other words, any acyclic connected graph is a tree. (Think about why the mentioned graph is a tree)

5. Suppose that we have two random variables named X and Y, these variables can have random values between 1 to 20, compute the covariance of X + Y and X - Y, are X +Y and X - Y independent? Explain your answer.

6. Suppose that we have three server computers, X, Y, Z. They receive requests independently. The moment-generating functions for the request distributions of the server computers are:

$$M_X = e^{3(e^t-1)}$$
$$M_Y = e^{5(e^t-1)}$$
$$M_Z = e^{(e^t-1)}$$

   Let J denotes sum of the requests for the servers. calculate $E[J^3]$.

R Problem

In many real world problems, scientists want to know if there is any interaction between two objects. Biologists want to know if two function of two genes are related. Economists tend to find features of a good market by measuring relation between features and the results. Covariance and correlation are measures of interaction between two random variables and how each one is dependent on the other.

Exploring IMDB +5000 Film Dataset

The supplementary file data.csv has been downloaded from kaggle and contains the data extracted from more than 5000 films on IMDB.

1. Write a code in R to calculate the covariance matrix for the following features and write the resulting matrix in 'cov_mat.csv':
   budget
   color
   num_critic_for_reviews
   duration
   director_facebook_likes

actor_1_facebook_likes
num_voted_users
cast_total_facebook_likes
num_user_for_reviews
imdb_score
movie_facebook_likes

2. Report your interpretation, which features are important for directing a good film?

3. Visualize the results for the features in a, pairwise, in an Scatter plot.