Statistical and Mathematical Methods DS-GA 1002   September 28, 2015

Homework 3
Due Tuesday, October 6

Please either give the assignment to Loraine at the CDS or send it via email
to the graders **before noon**.

1. *Messenger* Let us model Manhattan as a 13 x 2 miles rectangle. A package delivery company is situated right at the center.

   a. Assuming that the delivery points are uniformly spread in all of Manhattan, what is the average distance the messengers have to travel to deliver the packages? Take into account that they cannot travel in a straight line as there are buildings in between, so for a point (x, y) the distance is $|x| + |y|$.

   b. Upper bound the probability that a messenger has to travel more than 5 miles using Markov's inequality.

2. *Pasta and rice* You are hired by the management of a restaurant to model its stock probabilistically. You talk to the cook and he says:

   We cook pasta and rice. We always make sure that we have at least 100 lb of pasta or 100 lb of rice (if there is at least 100 lb of pasta, for example, we could have no rice at all); the logic being that we have daily specials and we want to be able to feed a lot of people with the same dish. However we never have more than 300 lb of rice or of pasta because we have no space to store it (we are able to store 300 lb of rice and 300 lb of pasta at the same time).

   You decide to model the quantity of pasta as a random variable X and the quantity of rice as a random variable R. As you have no information beyond what you have heard, so you assume that their joint pdf is constant (within the restrictions that you deduce from talking to the cook).

   a. Draw the joint pdf of X and R.

   b. Are X and R uncorrelated? Justify your answer. (Hint: You may want to use symmetry arguments to argue that some integrals equal zero without computing them.)

   c. Are X and R independent? Justify your answer.

3. *Restaurant* The management at the restaurant tells you:

   On average we have 40 customers every night and on average each customer spends 30 dollars, so on average we make 1200 dollars per night.

   a. Under what assumption is this true? Model the problem probabilistically and justify your answer.

   To explain that what they are saying is not necessarily the case, you tell them:

   Imagine that you only have either good nights in which you get 100 customers or bad nights in which you get 10 customers. On good nights people get annoyed by how crowded the restaurant is and they spend on average 10 dollars each. On bad nights the customers are relaxed and tend to drink more wine so they each spend 40 dollars on average.

   b. What is the probability of having a good night so that the average number of customers is 40?

c. In this imagined scenario, what is the average amount that each customer spends per night?

d. What is the average money that the restaurant makes under these assumptions? Why are you telling this story to the management?

4. *Copper* A company which buys and sells copper has just lost most of their data. They come to you and ask:

We only know that on average we have had 2 million dollars worth of copper in stock over the last ten years. In that span the price of copper has not gone above 5 dollars/lb and we cannot keep more than half a million pounds because of space restrictions. How likely is it that at any point we have had less than 1 million dollars worth of copper in stock?

a. Provide an upper bound for the probability that the company is interested in.

b. Is it sensible to model the price of copper and the amount of copper in stock as independent random variables?

c. Whatever your answer in (b), you decide to model the price and the amount as independent random variables. You look up the mean and the standard deviation of the price, which is 4.5 dollars/lb (mean) and 0.2 dollars/lb (standard deviation), and you learn from the company that the standard deviation of stored copper has been of 10 000 lbs. Compute another bound for the probability of interest under your assumptions using this information.

5. *Law of conditional variance* Let us define the conditional variance in a similar way to the conditional expectation.

a. What is the object $\mathrm{Var}\,(Y|X=x)$ (i.e. is it a number, a random variable or a function)? What does it represent?

b. Setting $h(x) = \mathrm{Var}\,(Y|X=x)$ we define the conditional variance as $\mathrm{Var}\,(Y|X) = h(X)$. What is this object?

c. Prove the law of conditional variance:

$$\mathrm{Var}\,(Y) = \mathrm{E}\,(\mathrm{Var}\,(Y|X)) + \mathrm{Var}\,(\mathrm{E}\,(Y|X)) \tag{1}$$

and describe it in words.

d. We model the time at which a runner gets injured (in hours) during a marathon as an exponential random variable with parameter equal to 1 if the runner is under 30 years old and 2 if she is over 30. What is the mean and the standard deviation of the time at which a runner gets injured if 20% of the runners are over 30?