Stat 230

Homework Problem Set 6

Due Jun 16 7:00 pm EST.

Please justify your answers and show the steps that lead you to your answer. Without a proper explanation, even a perfectly correct answer will receive a low score. Doubts regarding the problem sets/ Problem setup can be posted on Piazza. Simplify the expressions whenever you can.

Pr. 1 (15 pts)

Let $X_1, \ldots, X_n$ be random variables with variance 1 each, and let the correlation between any pair of distinct X_i, X_j be $1/4$.

Find $\mathrm{Var}(X_1 + X_2 + \ldots + X_n)$.

Pr. 2 [Erdös-Renyi Graphs] (15 pts)

The Erdös-Renyi Machine is a simple probabilistic model for graphs that has been used to prove existence of graphs with certain properties[1]. This took the mathematics community by surprise as the consensus was this is not possible - constructing such graphs is rather hard. The Erdös-Renyi model is still used today in network analysis problems. This kind of thinking inspired one of the most beautiful results in modern mathematics - *Szemeredi regularity lemma* that has far reaching consequences in arithmetic sequences and ergodic theory and that won Szemeredi the Abel Prize in 2012.

The Erdös-Renyi model is a distribution over graphs, and it has two parameters n and p. Let X be a graph on n vertices $X_1, \ldots, X_n$. For each pair of distinct vertices X_i, X_j, connect them with probability p, independently of other connections. The distribution of graphs constructed this way is the *Erdös-Renyi graph model* with parameters n and p, and for X generated by the process above, we write $X \sim \mathrm{ER}(n, p)$.

a) (5 pts) What is the distribution of number of overall connections in an Erdös-Renyi graph on n vertices and connection probability p?

b) (5 pts) Let D_k be the number of connections to a specific vertex X_k. What is the distribution of the connections D_k to vertex X_k?

c) (5 pts) Find the correlation between $D_i, D_j, i \neq j$.

Pr. 3 (20 pts)

a) (7 pts) Show that if $Z \sim \mathrm{Normal}(0, 1)$, then $Z^2 \sim \mathrm{Gamma}(1/2, 1/2)$. [Hint: use the mgf]

[1] E.g. graphs with simultaneously high girth and large chromatic number

b) (7 pts) Suppose (X, Y) are the coordinates of a random point Q on the plane, where $X \sim \text{Normal}(0,1)$,$Y \sim \text{Normal}(0,1)$ are independent. Let $R^2 = X^2 + Y^2$, so that R denotes the distance of Q from the origin. Show that $R^2 \sim \text{Exponential}(1/2)$. Calculate $\mu := \mathbb{E}(R^2)$. Also calculate the probability $\Pr(R \leq 0.5 \cdot \sqrt{\mu})$.

c) (6 pts) Now suppose Q is a random point on the 50-dimensional Euclidean space, with coordinates $(X_1, X_2, \ldots, X_{50})$, where each $X_j \sim \text{Normal}(0,1)$ and they are mutually independent. Again, let $R^2 = X_1^2 + \ldots + X_{50}^2$. Calculate $\mu := E(R^2)$. Also calculate the probability $\Pr(R \leq 0.5 \cdot \sqrt{\mu})$. [Hint: recognize that R^2 is the sum of fifty independent random variables.]

Pr. 4 (25 pts) Suppose that internet users access a particular Web site according to a Poisson process with a rate of λ per hour, but λ is unknown. The admin believes that $\lambda \sim \text{Exponential}(2)$. Let X be the number of users who access the Web site during a one-hour period.

a) (15 pts) Show that the marginal distribution of X is a geometric distribution. Determine the value of the parameter of the distribution. *Hint: Use the tower rule, LOTUS, and the moment generating function.*

b) (10 pts) If $X = 2$ is observed, show that the conditional distribution of λ, that is $p(\lambda|X = 2)$, is a gamma distribution. State the parameters of this distribution.

Pr. 5 (25 pts) Let Z be the number of cars that arrive at a busy intersection between 7:00 AM and 7:01 AM on a Monday morning. Let X denote the number of SUVs among these cars, and $Y := Z - X$ the number of non-SUVs. Suppose we believe that Z is distributed according to Poisson(100). Assume there is a 40% chance that any car arriving at the intersection is an SUV.

a) (5 pts) Argue that $X|(Z = z) \sim \text{Binomial}(z, 0.4)$ for each $z = 1, 2, \ldots$, and give an expression of the mgf of X given $Z = z$. [Hint: Let $X_i = 1$ if the ith car is an SUV and 0 otherwise.]

b) (7 pts) Derive the mgf of X by using the tower rule and your answer from part (a)

c) (3 pts) Show that $X \sim \text{Poisson}(40)$

d) (2 pts) Show that $Y \sim \text{Poisson}(60)$

e) (5 pts) Are X and Y independent? Why or why not? Justify your answer with a proof.

f) (3 pts) Let Z', X', Y' be independent Poisson random variables with parameters 100, 40 and 60, respectively. Is the distribution of (X, Y, Z) same as (X', Y', Z')? Why or why not? You don't need to prove anything, just state your answer with a justification.