UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 2

Spring 2019

---

1. **Poisson Merging**

   The Poisson distribution is used to model *rare events*, such as the number of customers who enter a store in the next hour. The theoretical justification for this modeling assumption is that the limit of the binomial distribution, as the number of trials n goes to ∞ and the probability of success per trial p goes to 0, such that $np \to \lambda > 0$, is the Poisson distribution with mean λ.

   Now, suppose we have two independent streams of rare events (for instance, the number of female customers and male customers entering a store), and we do not care to distinguish between the two types of rare events. Can the combined stream of events be modeled as a Poisson distribution?

   Mathematically, let X and Y be independent Poisson random variables with means λ and μ respectively. Prove that $X + Y \sim \text{Poisson}(\lambda + \mu)$. (This is known as **Poisson merging**.) Note that it is **not** sufficient to use linearity of expectation to say that $X + Y$ has mean $\lambda + \mu$. You are asked to prove that the *distribution* of $X + Y$ is Poisson.

   Note: This property will be extensively used when we discuss Poisson processes.

2. **Sampling without Replacement**

   Suppose you have N items, G of which are good and B of which are bad (B, G, and N are positive integers, $B + G = N$). You start to draw items without replacement, and suppose that the first good item appears on draw X. Compute the mean and variance of X.

3. **Clustering Coefficient**

   This problem will explore an important probabilistic concept of clustering that is widely used in machine learning applications today. Consider n students, where n is a positive integer. For each pair of students $i, j \in \{1, \dots, n\}$, $i \neq j$, they are friends with probability p, independently of other pairs. We assume that friendship is mutual. We can see that the friendship among the n students can be represented by an undirected graph G. Let $N(i)$ be the number of friends of student i and $T(i)$ be the number of triangles attached to student i. We define the **clustering coefficient** $C(i)$ for student i as follows:

   $$C(i) = \frac{T(i)}{\binom{N(i)}{2}}.$$

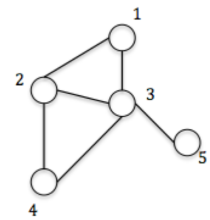

Figure 1: Friendship and clustering coefficient.

The clustering coefficient is not defined for the students who have no friends. An example is shown in Figure 1. Student 3 has 4 friends (1, 2, 4, 5) and there are two triangles attached to student 3, i.e., triangle 1-2-3 and triangle 2-3-4. Therefore $C(3) = 2/\binom{4}{2} = 1/3$.

Find $\mathbb{E}[C(i) \mid N(i) \geq 2]$.