UC Berkeley
Department of Electrical Engineering and Computer Sciences

Electrical Engineering 126: Probability and Random Processes

Discussion 2
Spring 2017

---

1. **Poisson Properties**

   (a) Suppose X and Y are independent Poisson random variables with mean λ and μ respectively. Prove that $X+Y$ has the Poisson distribution with mean $\lambda+\mu$. (This is known as **Poisson merging**.)

   (b) Suppose that X has the Poisson distribution with mean λ. View X as the number of arrivals of a process. Independently, for each arrival, mark the arrival as 0 with probability p and 1 with probability $1-p$. Let Y be the number of 0 arrivals and Z be the number of 1 arrivals. Prove that Y and Z are independent Poisson random variables with means λp and $\lambda(1-p)$ respectively. (This is known as **Poisson splitting**.)

Solution:

(a)

$$
\begin{aligned}
\Pr(X+Y=z) &= \sum_{j=0}^{z} \Pr(X=j, Y=z-j) = \sum_{j=0}^{z} \frac{e^{-\lambda}\lambda^j}{j!}\frac{e^{-\mu}\mu^{z-j}}{(z-j)!} \\
&= \frac{e^{-(\lambda+\mu)}}{z!}\sum_{j=0}^{z}\frac{z!}{j!(z-j)!}\lambda^j\mu^{z-j} \\
&= \frac{e^{-(\lambda+\mu)}}{z!}\sum_{j=0}^{z}\binom{z}{j}\lambda^j\mu^{z-j} = \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}
\end{aligned}
$$

(b) We prove that Y has the Poisson distribution with mean λp. The proof for the distribution of Z is similar.

$$
\begin{aligned}
\Pr(Y=y) &= \sum_{x=y}^{\infty}\Pr(X=x, Y=y) = \sum_{x=y}^{\infty}\Pr(X=x)\Pr(Y=y \mid X=x) \\
&= \sum_{x=y}^{\infty}\frac{e^{-\lambda}\lambda^x}{x!}\binom{x}{y}p^y(1-p)^{x-y} \\
&= e^{-\lambda}\sum_{x=y}^{\infty}\frac{\lambda^x}{x!}\frac{x!}{y!(x-y)!}p^y(1-p)^{x-y} \\
&= \frac{e^{-\lambda}(\lambda p)^y}{y!}\sum_{x=y}^{\infty}\frac{(\lambda(1-p))^{x-y}}{(x-y)!} = \frac{e^{-\lambda}(\lambda p)^y}{y!}e^{\lambda(1-p)} \\
&= \frac{e^{-\lambda p}(\lambda p)^y}{y!}
\end{aligned}
$$

Next, we prove that Y and Z are independent.

$$\begin{aligned}
\Pr(Y=y, Z=z) &= \sum_{x=0}^{\infty} \Pr(X=x, Y=y, Z=z) \\
&= \sum_{x=0}^{\infty} \Pr(Y=y, Z=z \mid X=x) \Pr(X=x) \\
&= \Pr(Y=y, Z=z \mid X=y+z) \Pr(X=y+z) \\
&= \frac{\cancel{(y+z)!}}{y!z!} p^y (1-p)^z \frac{e^{-\lambda}\lambda^{y+z}}{\cancel{(y+z)!}} \\
&= \frac{e^{-\lambda p}(\lambda p)^y}{y!} \cdot \frac{e^{-\lambda(1-p)}(\lambda(1-p))^z}{z!}
\end{aligned}$$

Remark: These properties will be used extensively when we discuss the Poisson process model.

2. **Sampling Without Replacement**

Suppose you have N items, G of which are good and B of which are bad ($B+G=N$). 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.

Solution:

The expectation is computed with a clever trick: let X_i be the indicator that the ith bad item appears before the first good item. Then, $X = 1 + \sum_{i=1}^{B} X_i$, and by linearity of expectation,

$$E[X] = 1 + BE[X_i] = 1 + \frac{B}{G+1} = \frac{N+1}{G+1}.$$

Observe that $\text{var}(X) = \text{var}(X-1)$. Using the same indicators, we compute $E[(X-1)^2]$.

$$\begin{aligned}
E[(X-1)^2] &= BE[X_i^2] + B(B-1)E[X_i X_j] \\
&= \frac{B}{G+1} + \frac{2B(B-1)}{(G+1)(G+2)}
\end{aligned}$$

Therefore, our answer is

$$\text{var}(X) = \frac{B}{G+1} + \frac{2B(B-1)}{(G+1)(G+2)} - \left(\frac{B}{G+1}\right)^2.$$

With a little algebra, we can actually simplify the result.

$$\begin{aligned}
\text{var}(X) &= \frac{B(G+1)(G+2) + 2B(B-1)(G+1) - B^2(G+2)}{(G+1)^2(G+2)} \\
&= \frac{BG(N+1)}{(G+1)^2(G+2)}
\end{aligned}$$

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. For each pair of students, say student i and student 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:

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

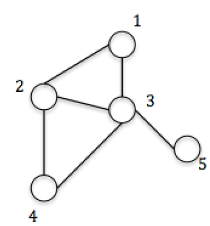

Figure 1: Friendship and clustering coefficient.

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 $E[C(i) \mid N(i) \geq 2]$.

Solution:

First, we compute $E[C(i) \mid N(i) = k]$, $(k \geq 2)$. Suppose that student i has friends $f_1, \ldots, f_k$. We can see that $T(i)$ equals the number of friend pairs among $\{f_1, \ldots, f_k\}$. Since there are $\binom{k}{2}$ possible pairs and each pair of students are friends with probability p, independently of other pairs, we know that the expected number of friend pairs among $\{f_1, \ldots, f_k\}$ is $\binom{k}{2}p$. Then we have

$$E[C(i) \mid N(i) = k] = \frac{\binom{k}{2}p}{\binom{k}{2}} = p.$$

Since this is true for all $k \geq 2$, we have $E[C(i) \mid N(i) \geq 2] = p$.

4. **Packet Routing**

Consider a system with n inputs and n outputs. At each input, a packet appears independently with probability p. If a packet appears, it is destined for one of the n outputs uniformly randomly, independently of the other packets.

(a) Let X denote the number of packets destined for the first output. What is the distribution of X?

(b) What is the probability of a collision, that is, more than one packet heading to the same output?

Solution:

(a) The probability that there exists a packet at an input and the packet is destined for the first output is p/n. By the independence over inputs, X has the binomial distribution $(n, p/n)$.

(b) Let C be the event of a collision and let N be the total number of packets in all inputs.

$$
\begin{aligned}
\Pr(C) = 1 - \Pr(\bar{C}) &= 1 - \sum_{k=0}^{\infty} \Pr(\bar{C} \mid N = k) \Pr(N = k) \\
&= 1 - \sum_{k=0}^{n} \frac{n!}{(n-k)! n^k} \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}
$$