UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 2
Fall 2021

1. Limit of Binomial

Show that the limit of a Binomial(n, p) distribution is Poisson(λ), where we take $n \to \infty$ and keep $\lambda = np$ fixed.

Solution: We write

$$\begin{aligned} P(\text{Binomial}(n,p) = k) &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)p^k(1-p)^{n-k}}{k!} \end{aligned}$$

Taking the limit as $n = \lambda/p \to \infty$, this becomes

$$\lim_{n\to\infty} \lambda^k (1-\lambda/n)^{n-k} \frac{1}{k!} = \frac{\lambda^k e^{-\lambda}}{k!},$$

where we have used the approximation

$$\lim_{n\to\infty} (1-\lambda/n)^n = e^{-\lambda}.$$

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.

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, for $i = 1, \dots, B$. Then, $X = 1 + \sum_{i=1}^{B} X_i$, and by linearity of expectation,

$$\mathbb{E}[X] = 1 + B\mathbb{E}[X_1] = 1 + \frac{B}{G+1} = \frac{N+1}{G+1}.$$

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

$$\begin{aligned} \mathbb{E}[(X-1)^2] &= B\mathbb{E}[X_1^2] + B(B-1)\mathbb{E}[X_1 X_2] \\ &= \frac{B}{G+1} + \frac{2B(B-1)}{(G+1)(G+2)} \end{aligned}$$

Therefore, our answer is

$$\operatorname{var} X = \frac{B}{G+1} + \frac{2B(B-1)}{(G+1)(G+2)} - \Big(\frac{B}{G+1}\Big)^2.$$

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

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

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, where n is a positive integer. For each pair of students $i, j \in \{1, \ldots, n\}$, $i \neq j$, 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:

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

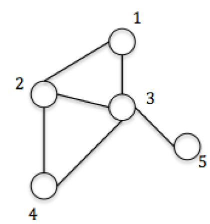

Figure 1: Friendship and clustering coefficient.

The clustering coefficient is not defined for the students who have no friends. An example is shown in Figure **??**. 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 $\mathbb{E}[C(i) \mid N(i) \geq 2]$.

Solution:

First, we compute $\mathbb{E}[C(i) \mid N(i) = k]$, for $k \in \{2, \ldots, n-1\}$. 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

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

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