Will Monroe
CS 109

Lecture Notes #8
July 12, 2017

The Poisson Distribution

Based on a chapter by Chris Piech

Binomial in the Limit

Recall the example of sending a bit string over a network. In our last class we used a binomial random variable to represent the number of bits corrupted out of 4 with a high corruption probability (each bit had independent probability of corruption $p = 0.1$). That example was relevant to sending data to spacecraft, but for earthly applications like HTML data, voice or video, bit streams are much longer (length $\approx 10^4$) and the probability of corruption of a particular bit is very small ($p \approx 10^{-6}$). Extreme n and p values arise in many cases: # visitors to a website, #server crashes in a giant data center.

Unfortunately, $X \sim \text{Bin}(10^4, 10^{-6})$ is unwieldy to compute. However, when values get that extreme, we can make approximations that are accurate and make computation feasible. Recall that the parameters of the binomial distribution are $n = 10^4$ and $p = 10^{-6}$. First, define $\lambda = np$. We can rewrite the binomial PMF as follows:

$$
\begin{aligned}
P(X = i) &= \frac{n!}{i!(n-1)!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1)\ldots(n-i-1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}
\end{aligned}
$$

This equation can be made simpler using some approximations that hold when n is sufficiently large and p is sufficiently small:

$$
\begin{aligned}
&\frac{n(n-1)\ldots(n-i-1)}{n^i} \approx 1 \\
&(1-\lambda/n)^n \approx e^{-\lambda} \\
&(1-\lambda/n)^i \approx 1
\end{aligned}
$$

Using these reduces our original equation to:

$$
P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}
$$

This simplification, derived by assuming extreme values of n and p, turns out to be so useful that it gets its own random variable type: the **Poisson random variable**.

Poisson Random Variable

A Poisson random variable approximates Binomial where n is large, p is small, and $\lambda = np$ is "moderate". Interestingly, to calculate the things we care about (PMF, expectation, variance), we no longer need to know n and p. We only need to provide λ, which we call the **rate**.

There are different interpretations of "moderate". Commonly accepted ranges are $n > 20$ and $p < 0.05$ or $n > 100$ and $p < 0.1$.

Here are the key formulas you need to know for Poisson. If Y is a Poisson random variable, denoted $Y \sim \text{Poi}(\lambda)$, then:

$$\begin{aligned} P(Y = i) &= \frac{\lambda^i}{i!} e^{-\lambda} \\ E[Y] &= \lambda \\ \text{Var}(Y) &= \lambda \end{aligned}$$

Example 1

Let's say you want to send a bit string of length $n = 10^4$ where each bit is independently corrupted with $p = 10^{-6}$. What is the probability that the message will arrive uncorrupted? You can solve this using a Poisson with $\lambda = np = 10^4 10^{-6} = 0.01$. Let $X \sim \text{Poi}(0.01)$ be the number of corrupted bits. Using the PMF for Poisson:

$$\begin{aligned} P(X = 0) &= \frac{\lambda^i}{i!} e^{-\lambda} \\ &= \frac{0.01^0}{0!} e^{-0.01} \\ &\approx 0.9900498 \end{aligned}$$

We could have also modeled X as a binomial such that $X \sim \text{Bin}(10^4, 10^{-6})$. That would have been harder to compute but would have resulted in the same number (to 8 decimal places).

Example 2

The Poisson distribution is often used to model the number of events that occur independently at any time in an interval of time or space, with a constant average rate. Earthquakes are a good example of this. Suppose there are an average of 2.8 major earthquakes in the world each year. What is the probability of getting more than one major earthquake next year?

Let $X \sim \text{Poi}(2.8)$ be the number of major earthquakes next year. We want to know $P(X > 1)$. We can use the complement rule to rewrite this as $1 - P(X = 0) - P(X = 1)$. Using the PMF for Poisson:

$$\begin{aligned} P(X > 1) &= P(X = 0) - P(X = 1) \\ &= 1 - e^{-2.8} \frac{2.8^0}{0!} - e^{-2.8} \frac{2.8^1}{1!} \\ &= 1 - e^{-2.8} - 2.8e^{-2.8} \\ &\approx 1 - 0.06 - 0.17 \\ &= 0.77 \end{aligned}$$