Lisa Yan and Jerry Cain
CS109

Lecture Notes #8
September 30, 2020

The Poisson Distribution and other Discrete Distributions

Based on a chapter by Chris Piech

Pre-recorded lecture: Section 1 through Section 4.
In-lecture: Review, focusing on Poisson approximation of Binomial.
Not covered: Section 5

1 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:

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

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**.

2 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:

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

3 Geometric Distribution

X is a **geometric random variable** ($X \sim \text{Geo}(p)$) if X is number of the independent trials until the first success and p is probability of success on each trial. If $X \sim \text{Geo}(p)$:

$$\begin{aligned} P(X = n) &= (1-p)^{n-1}p \\ E[X] &= 1/p \\ \text{Var}(X) &= (1-p)/p^2 \end{aligned}$$

The PMF, $P(X = n)$, can be derived using the independence assumption. Let E_i represent the event that the i-th trial succeeds. Then the probability that X is exactly n is the probability that the first $n-1$ trials fail, and the n-th succeeds:

$$\begin{aligned} P(X = n) &= P({E_1}^C {E_2}^C \dots {E_{n-1}}^C E_n) \\ &= P({E_1}^C)P({E_2}^C)\dots P({E_{n-1}}^C)P(E_n) \\ &= (1-p)^{n-1}p \end{aligned}$$

A similar argument can be used to derive the CDF, the probability that $X \leq n$. This is equal to $1 - P(X > n)$, and $P(X > n)$ is the probability that at least the first n trials fail:

$$\begin{aligned} P(X \leq n) &= 1 - P(X > n) \\ &= 1 - P({E_1}^C {E_2}^C \dots {E_n}^C) \\ &= 1 - P({E_1}^C)P({E_2}^C)\dots P({E_n}^C) \\ &= 1 - (1-p)^n \end{aligned}$$

Example 1

In the *Pokémon* games, one captures Pokémon by throwing Poké Balls at them. Suppose each ball independently has probability $p = 0.1$ of catching the Pokémon.

Problem: What is the average number of balls required for a successful capture?

Solution: Let X be the number of balls used until (and including) the capture. $X \sim \text{Geo}(p)$, so the average number needed is $E[X] = 1/p = 10$.

Problem: Suppose we want to ensure that the probability of a capture before we run out of Poké Balls is at least 0.99. How many balls do we need to carry?

Solution: We want to know n such that $P(X \leq n) \geq 0.99$.

$$\begin{aligned} P(X \leq n) = 1 - (1-p)^n &\geq 0.99 \\ (1-p)^n &\leq 0.01 \\ \log[(1-p)^n] &\leq \log 0.01 \\ n \log(1-p) &\leq \log 0.01 \\ n &\geq \frac{\log 0.01}{\log(1-p)} = \frac{\log 0.01}{\log 0.9} \approx 43.7 \end{aligned}$$

So we need 44 Poké Balls. (Note that we flipped the inequality on the last line because we divided both sides by $\log(1-p)$. Since $1 - p < 1$, we know $\log(1-p) < 0$, so we're dividing by a negative number!)

4 Negative Binomial Distribution

X is a **negative binomial random variable** ($X \sim \text{NegBin}(r, p)$) if X is the number of independent trials until r successes and p is probability of success on each trial. If $X \sim \text{NegBin}(r, p)$:

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \text{ where } r \leq n$$
$$E[X] = r/p$$
$$\text{Var}(X) = r(1-p)/p^2$$

Example 2

Problem: A grad student needs 3 published papers to graduate. (Not how it works in real life!) On average, how many papers will the student need to submit to a conference, if the conference accepts each paper randomly and independently with probability $p = 0.25$? (Also not how it works in real life...though the NIPS Experiment[1] suggests there is a grain of truth in this model!)

Solution: Let X be the number of submissions required to get 3 acceptances. $X \sim \text{NegBin}(r = 3, p = 0.25)$. So $E[X] = \frac{r}{p} = \frac{3}{0.25} = 12$.

5 Other distributions

Hypergeometric Distribution

The remaining three distributions appear occasionally; you don't have to master them for this course, but it can be useful to know they exist.

X is a **hypergeometric random variable** ($X \sim \text{HypG}(n, N, m)$) if X is the number of red balls drawn when n balls are drawn at random, *without replacement*, from an urn with N balls total, m of which are red. If $X \sim \text{HypG}(p)$:

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \text{ where } 0 \leq k \leq \min(n, m)$$
$$E[X] = n\frac{m}{N}$$
$$\text{Var}(X) = \frac{nm(N-n)(N-m)}{N^2(N-1)}$$

[1] http://blog.mrtz.org/2014/12/15/the-nips-experiment.html

Benford Distribution

Benford's law says that "naturally occurring" numbers have an uneven distribution of their *first digits*. This is because, roughly speaking, many collections of numbers are not evenly distributed, but rather their *logs* are evenly distributed. The law says that the fraction of numbers with a first digit of 1 is usually close to $\log_{10}\left(1+\frac{1}{1}\right) \approx 0.301$, the fraction with a first digit of 2 is close to $\log_{10}\left(1+\frac{1}{2}\right) \approx 0.176$, and so on. This forms a probability distribution over the numbers 1 through 9.

More generally, in number base b (for example, in hexadecimal $b = 16$), X is distributed according to Benford's law if:

$$P(X = d) = \log_b\left(1 + \frac{1}{d}\right) \text{ where } 1 \leq d < b$$
$$E[X] = (b-1) - \log_b[(b-1)!]$$

Zipf Distribution

X is a **Zipf random variable** ($X \sim \text{Zipf}(s, N)$) if the probability of X obeys an *inverse power law*:

$$P(X = k) = C \cdot \frac{1}{k^s} \text{ where } 1 \leq k \leq N$$

where C is a normalizing constant (which turns out to be equal to reciprocal of the Nth harmonic number).

In human languages, a Zipf distribution is a good model of the frequency rank index of a randomly chosen word, where N is the number of words in the language, and s also depends on various properties of the language (but is often close to 1). Other processes involving rank-ordering quantities also frequently result in a Zipf distribution, such as the rank of populations of large cities.