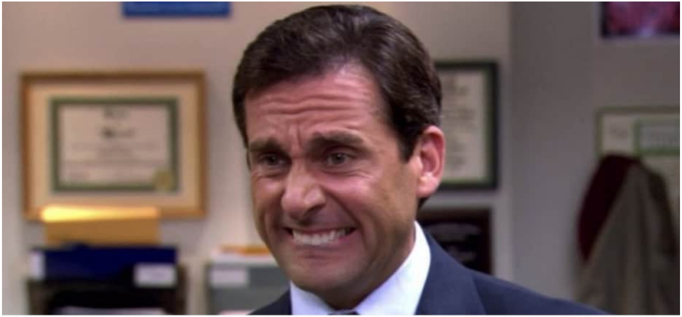
STAT 230 SECTION 2 LECTURE 14



When you realize your instructor needs to quarantine and you will only have fun online lectures instead of fun in person lectures

Today's Agenda

Last time:

- Binomial Distribution (Section 5.4)
- Negative Binomial Distribution (Section 5.5)

Today (Lec 14, 06/06):

- Geometric Distribution (Section 5.6)
- Memoryless property of the geometric distribution
- Poisson distribution (Section 5.7)

Negative Binomial Distribution

Definition

- Assume an experiment has two possible outcomes, "S" (success) and "F" (failure).
- Assume that in every try, $P(S) = p$ and $P(failure) = 1 - p$ for $0 < p < 1$.
- Repeat the experiment independently until a specified number k of successes have been observed.
- The number of failures before the kth success follows a Negative Binomial distribution, $X \sim NegBin(k, p)$.

The probability function of $X \sim NegBin(k, p)$ is

$$f(x) = P(X = x) = \binom{x+k-1}{x} p^k (1-p)^x, \quad x = 0, 1, 2, \ldots$$

Binomial versus Negative Binomial

Binomial distribution: We know number of trials n, but we do not know how many successes.

Negative Binomial distribution: We know the number of successes k, but we do not know how many trials will be needed.

...and because it's cool, we call the negative binomial distribution with $k = 1$ ("number of failures until the first success") the geometric distribution, notation $Geo(p)$.

Probability function and cumulative distribution function of the geometric distribution

- If $X \sim Geo(p)$, then X has probability function

$$f(x) = P(X = x) = (1-p)^x p, \quad x \in \{0, 1, 2, \dots\}.$$

- The cdf can be computed as

$$F(x) = P(X \leq x) = P(X \leq \lfloor x \rfloor) = \sum_{k=0}^{\lfloor x \rfloor} (1-p)^k p = 1 - (1-p)^{\lfloor x \rfloor + 1}$$

  if $x \geq 0$ and 0 otherwise.
- Note that $P(X > x) = 1 - F(x) = (1-p)^{\lfloor x \rfloor + 1}$ which is nice for computations!

Question

Dwight is flipping coins until the first head appears, each with their own coin. After 2 tosses, neither of them have seen their first head yet. Of course, he immediately runs to Michael to tell him.Dwight continuous flipping, and another 2 trials forgo without a head appearing. Michael Scott enters the room and says that this can't be correct! He argues that given that he needed more than 2 coin tosses, the probability he needs a total of more than 4 coin tosses until first head appears should be a lot smaller than the probability he needs more than 2 coin tosses because if we already had 2 fails, the coin must be heads soon!

Help them out!

Michael is, of course, wrong!

- If X denotes the number of failures until first head, we know $X \sim Geo(1/2)$. You found

$$P(X > 4 \mid X > 2) = \frac{P(X > 4)}{P(X > 2)} = P(X > 2)$$

- Reason: Due to independence of the trials, the coin forgets what it was flipped

Memoryless property

Example (Memoryless property of $Geo(p)$)

Let $X \sim Geo(p)$ and s, t be non-negative integers. Then, the following equation holds.

$$P(X \geq s + t | X \geq s) = P(X \geq t).$$

- Prove this!
- One can show that the geometric distribution is the only discrete distribution with this property.

Poisson distribution

Figure: Siméon Denis Poisson, 1781-1840

Poisson distribution

Definition

We say the random variable X has a **Poisson** distribution with parameter $\mu > 0$ if

$$f(x) = e^{-\mu}\frac{\mu^x}{x!}, \quad x = 0, 1, 2, \ldots$$

We denote this relationship as

$$X \sim Poisson(\mu) \text{ or } Poi(\mu),$$

and μ is often referred to as the "rate" parameter.

Exercise

Verify that f is a valid probability function!

Solution: Obviously, $f(x) \geq 0$ for all x, and

$$\begin{aligned}
\sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} e^{-\mu} \frac{\mu^x}{x!} \\
&= e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \\
&= e^{-\mu} e^{\mu} \\
&= 1.
\end{aligned}$$

The poisson distribution can easily be applied in real life situations to model the number of events that occurred, when events occur sequentially, and one occurrence is independent of the next.

a) Number of calls at a call centre
b) Number of car crashes at an intersection
c) Number of faulty iPhones produced in a factory

Poisson from Binomial

One way to view the Poisson distribution is to consider the limiting case of binomial distribution, where you fix $\mu = np$, and let $n \to \infty$ and $p \to 0$.

Example (Poisson approximation of Binomial)

Let $\mu = np$. Then if n is large and p is close to zero,

$$\binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\mu} \frac{\mu^x}{x!}$$

More precisely, one can show that if $n \to \infty$ and $p = p_n \to 0$ as $n \to \infty$ in such a way that $np_n \to \mu$, then

$$\binom{n}{x} p^x (1-p)^{n-x} \to e^{-\mu} \frac{\mu^x}{x!}, \quad as \quad n \to \infty$$

See course notes for the derivation!

Example

A bit error occurs for a given data transmission method independently in one out of every 1000 bits transferred. Suppose a 64 bit message is sent using the transmission system.

a) What is the probability that there are exactly 2 bit errors?

b) Approximate the probability in a) using a Poisson approximation.

Solution

a) X =number of errors, then $X \sim Bin(n = 64, p = 1/1000)$. We find

$$P(X = 2) = \binom{64}{2} \left(\frac{1}{1000}\right)^2 \left(\frac{999}{1000}\right)^{64-2} \approx 0.00189.$$

b) n is large, p is small, so X follows approximately a Poisson with $\mu = np = 64/1000$, so

$$P(X = 2) = e^{-\frac{64}{1000}} \frac{\left(\frac{64}{1000}\right)^2}{2!} \approx 0.00192.$$

The probabilities are almost the same!

Homework: Course notes up to and including 5.7

Next lecture: Finish Chapter 5. Feel free to read ahead!