Math 183
Statistical Methods

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

Today: Chapter 3 (continued)

- Negative Binomial Model
- Poisson Model
- Practice these models in different situations

Recap of Last Lecture

Geometric Model: $X = Geom(p)$

What is the probability that it takes exactly k Bernoulli trials to get the first success?
On average, how many Bernoulli trials will it take to get the first success?

Binomial Model: $X = Binom(n, p)$

What is the probability of getting exactly k successes in n Bernoulli trials?
On average, how many successes will I have in n Bernoulli trials?

Yet Another Common Question

In general, some behavior is average. How likely am I to see some specific behavior?

Examples:

Average behavior	Specific behavior
12.5 emails/day	0 email in one day
2.5 goals/game	9 goals in a game
1092 forest fires/year	1000 forest fires in a year

The **Poisson Model** is useful when you know the average behavior and want to explore specific cases. It requires two key ideas:

- An average value λ, known as the frequency parameter (12.5,2.5,1092)
- Some fixed time (day,game,year) in which you count something (emails,goals,fires).

The Poisson Model

The **Poisson Model** need a parameter $\lambda > 0$. Its distribution is, for $k = 0, 1, 2, \ldots$

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

where $k! = 1 \times 2 \times \ldots \times k$ (and $0! = 1$).
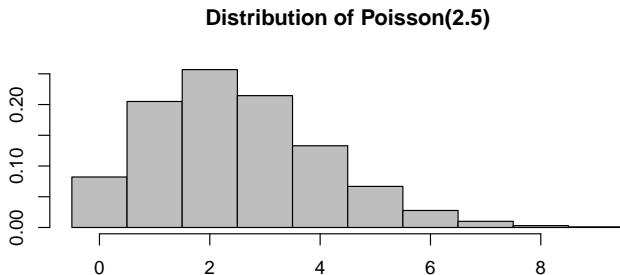
Notation: $X = Poisson(\lambda)$.

Example: The average number of goals in a soccer game is 2.5 goals/game. What is the probability of seeing a game with score 0-0?

We build a Poisson model with $\lambda = 2.5$: $X = Poisson(2.5)$. Then

$$P(X = 0) = e^{-2.5} \frac{2.5^0}{0!} \simeq 8.2\%.$$

The Poisson Model: Visualization

Take $X = Poisson(\lambda = 2.5)$.

Distribution of Poisson(2.5)

k	$P(X = k)$
0	$e^{-\lambda}$
1	$e^{-\lambda}\lambda$
2	$e^{-\lambda}\lambda^2/2$
$\ldots$	$\ldots$
k	$e^{-\lambda}\lambda^k/k!$
$\ldots$	$\ldots$

The Poisson Model: Parameters

If $X = Poisson(\lambda)$, one can show that

$$E(X) = \lambda \qquad\qquad Var(X) = \lambda.$$

Remark: the fact that $E(X) = \lambda$ is expected, since we built our model given an average behavior λ.

Example: With soccer games, $E(X) = 2.5$, hence $SD(X) = \sqrt{2.5} \simeq 1.58$ goals/match.
In other words, goals with goal totals between $2.5 - 1.58 = 0.92$ and $2.5 + 1.58 = 4.08$ are pretty common.

Common Question... Continued

Geometric Model: What is the probability of getting the <u>1st success</u> on the nth trial?

Negative Binomial Model: What is the probability of getting the <u>kth success</u> on the nth trial?

Remark:

- This should feel like Geometric, since it's about a number of trials before a certain number of successes. If we set $k = 1$, the negative binomial is just the geometric model.
- This should feel like Binomial, since it's about k sucesses in n trials. The difference is
    - for the Binomial, n is fixed, and you explore for different values of k.
    - for the Negative Binomial, k is fixed, and you explore the probability for different values of n.

Binomial VS Negative Binomial

Suppose there is a 20% chance a randomly-chose UCSD student comes from California.

Binomial question:
You talk to 7 students. What is the probability exactly 3 are from California?
$X = Binom(n = 7, p = 0.2)$ and $P(X = 3) = \binom{7}{3}0.2^3(1 - 0.2)^4$.

Negative Binomial question:
Your goals is to meet 3 students from California. What is the probability you'll meet the third one when you encounter the seventh person?
$Y = NegBinom(k = 3, p = 0.2)$ and $P(Y = 7) = ?$.

The Negative Binomial Model

The Negative Binomial Model says the following. If a Bernoulli trial has probability p of success, then the probability that the kth success will happen on the nth trial is, for $n \geq k$,

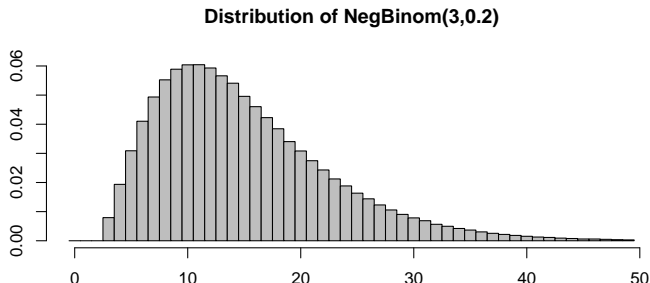$$P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}.$$

Notation: $NegBinom(k, p)$.

Example: $k = 3$, and explore $P(X = 5)$ ($n = 5$). The ways of getting the 3rd success on the 5th trial are:

SSFF**S**, SFSF**S**, SFFS**S**, FSSF**S**, FSFS**S**, FFSS**S**

- Each sequence must end with S. Therefore, the number of such sequences is the number of ways of arranging the other $(k - 1)$ S's in the remaining $(n - 1)$ spots, that is $\binom{n-1}{k-1}$.

- Each sequence contains $k = 3$ S's and $n - k = 2$ F's, so they have probability $p^k (1 - p)^{n-k}$.

The Negative Binomial Model: Visualization

Take $X = NegBinom(k = 3, p = 0.2)$.

Distribution of NegBinom(3,0.2)

n	$P(X = n)$
k	p^k
$k + 1$	$\binom{k}{k-1}p^k(1 - p)$
$k + 2$	$\binom{k+1}{k-1}p^k(1 - p)^2$
$\ldots$	$\ldots$
n	$\binom{n-1}{k-1}p^k(1 - p)^{n-k}$
$\ldots$	$\ldots$

The Negative Binomial Model: Parameters

If $X = NegBinom(k, p)$, one can show that

$$E(X) = \frac{k}{p} \qquad\qquad Var(X) = \frac{k(1-p)}{p^2}.$$

Remark: setting $k = 1$, you recognize formulas for the Geometric model.

Example: A basketball player scores a free throw with probability 0.2. After practice, he tries some over and over until he gets 10 that are in. On average, how many throws will he take to achieve this? How much variation might he expect from day to day?

$X = NegBinom(k = 10, p = 0.2)$, and we have

$$E(X) = \frac{10}{0.2} = 50 \text{ tries} \qquad\qquad SD(X) = \sqrt{\frac{10 \times (1 - 0.2)}{0.2^2}} \simeq 14.14 \text{ tries.}$$

So, total tries between 36 and 64 will be quite common.

Exercise

You flip an unfair coin ($p_{Heads} = 0.4$) until you get 5 Heads. What is the probability that the final Heads occurs of flip 6 or flip 7?

We have a fixed number of successes ($k = 5$) and are wondering about the number of trials this takes. This is a negative binomial:

$$X = NegBinom(k = 5, p = 0.4).$$

The questions asks

$$
\begin{aligned}
P(X = 6 \text{ or } X = 7) &= P(X = 6) + P(X = 7) \\
&= \binom{6-1}{5-1}(0.4)^5(1-0.4)^1 + \binom{7-1}{5-1}(0.4)^5(1-0.4)^2 \\
&\simeq 8.6\%.
\end{aligned}
$$

On average, how many flips will it take for you to get those 5 Heads?

$$E(X) = \frac{5}{0.4} = 12.5 \text{ flips.}$$

Summary of Our Models

Geometric: $X = Geom(p)$. $X \in \{1, 2, 3, \ldots\}$
X is the number of trials needed to get the first success. Each trial has success probability p.

Binomial: $X = Binom(n, p)$. $X \in \{0, 1, 2, \ldots, n\}$
X is the number of successful trials of out the number of trials. Each trial has success probability p.

Poisson: $X = Poisson(\lambda)$. $X \in \{0, 1, 2, \ldots\}$.
X is the number of times an event occurs in a given times when its average rate of occurrence in that time is λ.

Negative Binomial: $X = NegBinom(k, p)$. $X \in \{k, k+1, k+2, \ldots\}$
X is the number of trials needed to get the kth success. Each trial has success probability p.

Two Key Skills to Working Problems about Models

1. Deciding what model is the appropriate choice to describe the given situation
2. Deciding if the problem wants a probability, and expected value, or a standard deviation

Example: In a recent season, the UCSD women's wate polo team score 385 goals in 39 games. What's the probability they score 2 goals in a game?

We have an average rate in some time span: $385/39 \simeq 9.87$ goals/game. Poisson is appropriate: $X = Poisson(\lambda = 9.87)$.

$$P(X = 2) = e^{-9.87}\frac{9.87^2}{2!} \simeq 0.25\%.$$

How much variation is there is their point total per game?

$$SD(X) = \sqrt{9.87} \simeq 3.14 \text{ goals}.$$