Math 183
Statistical Methods

Eddie Aamari
S.E.W. Assistant Professor

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

Today: Chapter 3 (beginning)

- Bernoulli trials
- Geometric model
- Binomial model
- Expected value, variance and standard deviation of these models

Broader Models

Last lecture, we studied contrived situations and modeled them.

x	-4\$	72\$	-12\$
P(X = x)	1/2	$1/2 \times 1/6 = 1/12$	$1/2 \times 5/6 = 5/12$

Today, we are going to tackle common setups that arise frequently.

It'll lead us to standard models that are useful in many problems.

The Groundwork

Often, we have a random process that can result in **2 possible outcomes** and in which trials are **independent**.

Examples:
- Coin flip (heads, tails)
- Dice roll (even result, odd result)
- Gene mutated (yes, no)
- Email label (spam, not spam)
- Rugby game result (win, not win)

A **Bernoulli trial** is a random variable with precisely two outcomes and in which trials are independent.
You can call one a "success" and the other a "failure".

We usually write

x	Success	Failure
P(X = x)	p	$1 - p$

Exercise

The probability of my favorite rugby team winning a game is $p = 0.4$.
What is:

- the probability that my team loses its 4th game of the season?
  Answer: $1 - 0.4 = 0.6$

- the probability that they lose their first game and win their second?
  Answer: 0.6×0.4

- the probability that they lose their first two games and win the third?
  Answer: $(0.6)^2 \times 0.4$

- the probability that they lose every game in a 12-game season?
  Answer: $(0.6)^{12}$

A Common Question

What is the probability that it takes exactly k Bernoulli trials to get your first success?

On average, how many Bernoulli trials will it take to get the first success?

The probability model that answers questions about "first success" is called the **Geometric Model**.

The Geometric Model

Assume we are doing a Bernoulli trial with probability of success p (and failure $q = 1 - p$) over and over until success:

k	$P(X = k)$
1	p
2	qp
3	$q^2 p$
...	...
k	$q^{k-1} p$
...	...

Each row of the model answers a question like: "What is the probability that it takes k trials to get the first success?".

The Geometric Model: Parameters

The geometric model says that the probability of finally getting a success in k Bernoulli trials is

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

Notation: $X = Geom(p)$.
(p is known as a parameter, some value the model is built on).

One can show that
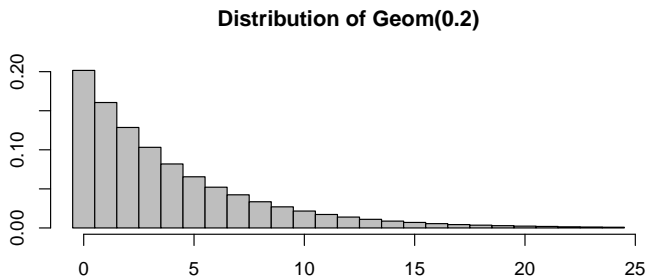
$$E(X) = \frac{1}{p} \qquad\qquad Var(X) = \frac{1-p}{p^2}.$$

Example: Your goal is to flip a coin until you get tails (= success). How many time, on average, do you expect this will take?

We have $X = Geom(0.5)$, so

$$E(X) = \frac{1}{0.5} = 2 \qquad\qquad SD(X) = \sqrt{\frac{1-0.5}{0.5^2}} = \sqrt{2}.$$

The Geometric Model: Visualization

Take $X = Geom(p = 0.2)$.

Distribution of Geom(0.2)

k	$P(X = k)$
1	p
2	$(1 - p)p$
3	$(1 - p)^2 p$
...	...
k	$(1 - p)^{k-1} p$
...	...

Practice

The probability that my favorite rugby team wins a game is 0.4. On average, how many games must they play until their first loss? Is their much variability?

We have a Geometric model with $p = 0.6$.
That is $X = Geom(0.6)$.

$$E(X) = \frac{1}{p} = \frac{1}{0.6} \simeq 1.66 \text{ games.}$$

$$SD(X) = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{0.4}{0.6^2}} \simeq 1.05 \text{ games.}$$

91% of all emails sent in the world are spam. As an experiment, you turn off your spam filter, and each day, you count how many emails you need to read before getting to an actual message.

What's the probability of reading three spam messages before getting an actual message?
We want

$P(\text{fail and fail and fail and success}) = 0.91 \times 0.91 \times 0.91 \times 0.09 \simeq 6.75\%.$

How many spams do you expect to get before having a genuine email?

$$E(X) = \frac{1}{p} = \frac{1}{0.09} \simeq 11.1 \text{ emails}$$

Another Common Question

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?

The probability model that answers questions about "how many of something in a fixed number of trials n" is called the **Binomial Model**.

Exercise

You flip a fair coin exactly 4 times ($n = 4$)

- What is the probability of no Heads ($k = 0$)?
  Answer: $(1/2)^4$

- Which is more likely: getting no Heads ($k = 0$), or getting 1 Head ($k = 1$)?
  Answer: There are 4 ways to get exactly 1 Head (HTTT,THTT,TTHT,TTTH). Each has probability $(1 - 1/2)^3 \times (1/2) = 1/16$. Adding all four gives $1/4$.

- What is the most likely outcome (most likely number of Heads)?
  Answer: $k = 2$ (getting 2 Heads). It has probability $6/16$.

The Binomial Model

Assume we conduct a Bernoulli trial (with success probability p and failure probability $q = 1 - p$) a total of n times.

k	$P(X = k)$
0	$q \times q \times \ldots \times q = q^n$
1	?
2	
$\ldots$	$\ldots$
n	$p \times p \times \ldots \times p = p^n$

For $k = 1$, there are many different ways to get 1 success:

$$SFFF\ldots, FSFFF\ldots, FFSFF\ldots, \ldots$$

Each of these has probability $q^{n-1}p$.

Binomial Coefficient (Choose Symbol)

The choose symbol helps you calculate how many ways there are to list one S among all those F's (the answer is n.)

It can also be used to count how many ways you could put <u>two</u> S's among all the F's.

In general, $\binom{n}{k}$ gives the count of how many ways you can get exactly k successes from n trials. The formula is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

where $k! = 1 \times 2 \times 3 \times \ldots \times k$.

R function: `choose(n,k)`

The Binomial Model

k	$P(X = k)$
0	$q \times q \times \ldots \times q = q^n$
1	$\binom{n}{1} q^{n-1} p$
2	$\binom{n}{2} q^{n-2} p^2$
...	...
n	$p \times p \times \ldots \times p = p^n$

Say $n = 5$ and $k = 2$. We have to take into account:

SSFFF, SFSFF, SFFSF, SFFFS, FSSFF, FSFSF, FSFFS, FFSSF, FFSFS, FFFSS,

which all have probability $q^3 p^2$.

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} \frac{1 \times 2 \times 3 \times 4 \times 5}{1 \times 2 \times 1 \times 2 \times 3} = \frac{4 \times 5}{2} = 10.$$

Remark: $\binom{n}{0} = \binom{n}{n} = 1.$

The Binomial Model

The **Binomial Model** says that the probability of getting k successes in n independent Bernoulli trials is

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

Notation: $X = Binom(n, p)$
(n is the number of trials, p is the success probability of each trial)

Example: What is the probability of getting 2 Heads when flipping a coin 7 times?

$n = 7$ and $p = 0.5$, so

$$P(X = 2) = \binom{7}{2} 0.5^2 (1-0.5)^5 \simeq 16.4\%.$$

The Binomial Model: Parameters

One can show that

$$E(X) = np \qquad \qquad Var(X) = np(1 - p).$$

Example: flipping a coin 7 times, how many Heads do we expect on average? Is there much variation in that average?
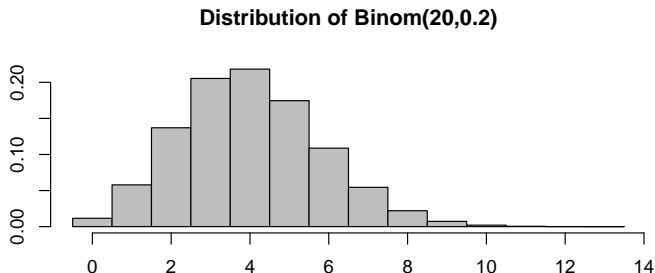
$$E(X) = 7 \times 0.5 = 3.5 \text{ Heads},$$

and

$$SD(X) = \sqrt{7 \times 0.5 \times (1 - 0.5)} \simeq 1.32 \text{ Heads}.$$

The Binomial Model: Parameters

Take $X = Binom(20, 0.2)$.

Distribution of Binom(20,0.2)

k	$P(X = k)$
0	$(1-p)^n$
1	$\binom{n}{1}p(1-p)^{n-1}$
2	$\binom{n}{2}p^2(1-p)^{n-2}$
...	...
n	p^n

Practice

The probability of my favorite rugby team winning a game is $p = 0.4$.

- On average, how many games do they have to play until they win a game?
  Answer: $X = Geom(0.4)$, and we want

$$E(X) = \frac{1}{0.4} = 2.5.$$

- In 15 games, what's the probability they win strictly more that 1 match?
  Answer: $Y = Binom(15, 0.4)$, and we want

$$\begin{aligned}
P(Y \geq 2) &= 1 - P(Y = 0) - P(Y = 1) \\
&= 1 - \binom{15}{0}(1 - 0.4)^{15} + \binom{15}{1}(1 - 0.4)^{14} \times 0.4^1 \\
&\simeq 99.48\%.
\end{aligned}$$

Announcement

Students on **Wait List**:
Check your emails TODAY and follow the instructions.
(mind spam box!)
If you got no email, you will not be enrolled in Math 183 for Fall.

Student in **Concurrent Enrollment**:
Come and see me to have your case cleared.

To **all**:

- Turn in Homework 2
- Read Chapter 3
- Practice Midterm I (Extra exercises available on the website)