Math 11
Calculus-Based Introductory Probability and Statistics

Eddie Aamari
S.E.W. Assistant Professor

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

Today:

- The common discrete probability 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	$q^2 p$
$\dots$	$\dots$
k	$q^{k-1} p$
$\dots$	$\dots$

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

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

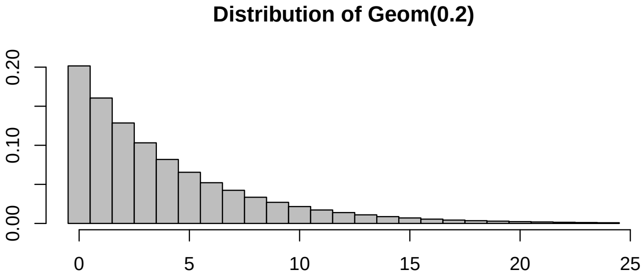
Example: Your goal is to flip a coin until you get tails (= success). How many time, on average, do you expect this will take?

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 $X = Geom(p = 0.2)$.



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

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 $X = Geom(0.6)$.

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

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

Practice

Which of the following ideas are Bernoulli trials?

1. What result comes up on a fair six-sided die
2. Whether each bite of a pizza tastes good or not
3. Whether a random number generator gives a result less than 1 or at least 1
4. Whether it rains or not each day this week

Answer: 3.

Practice

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	$\binom{n}{2} q^{n-2} p^2$
$\ldots$	$\ldots$
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

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

Example: flipping a coin 7 times, how many Heads do we expect on average? Is there much variation in that average?

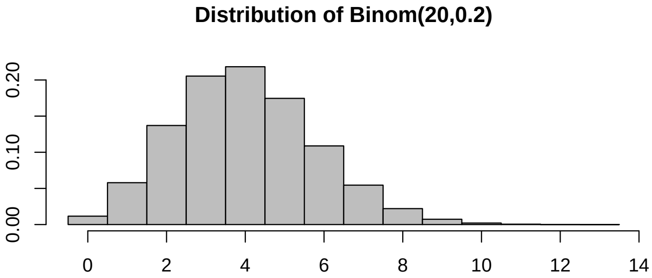
$$E(X) = 7 \times 0.5 = 3.5 \text{ Heads},$$

and

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

The Binomial Model: Parameters

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



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

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: $X = Poisson(\lambda)$.

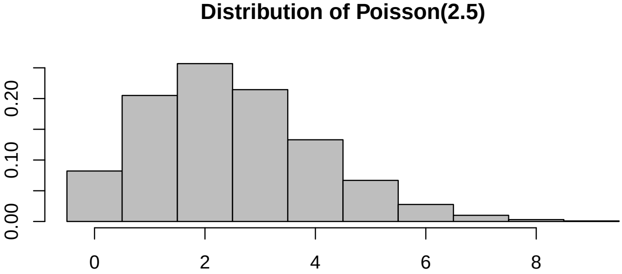
Example: The average number of goals in a soccer game is 2.5 goals/game. What is the probability of seeing a game with score 0-0?

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



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 reassuring, since we built our model given this 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.

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.

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: $385/39 \simeq 9.87$ goal-s/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?

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

Additional Practice

	% LGBT
San Francisco-Oakland-Hayward, Calif.	6.2
Portland-Vancouver-Hillsboro, Ore.-Wash.	5.4
Austin-Round Rock, Texas	5.3
New Orleans-Metairie, La.	5.1
Seattle-Tacoma-Bellevue, Wash.	4.8
Boston-Cambridge-Newton, Mass.-N.H.	4.8
Salt Lake City, Utah	4.7
Los Angeles-Long Beach-Anaheim, Calif.	4.6
Denver-Aurora-Lakewood, Colo.	4.6
Hartford-West Hartford-East Hartford, Conn.	4.6
Louisville/Jefferson County, Ky.-Ind.	4.5
Virginia Beach-Norfolk-Newport News, Va.-N.C.	4.4
Providence-Warwick, R.I.-Mass.	4.4
Las Vegas-Henderson-Paradise, Nev.	4.3
Columbus, Ohio	4.3
Jacksonville, Fla.	4.3
Miami-Fort Lauderdale-West Palm Beach, Fla.	4.2
Indianapolis-Carmel-Anderson, Ind.	4.2
Atlanta-Sandy Springs-Roswell, Ga.	4.2
Orlando-Kissimmee-Sanford, Fla.	4.1
Tampa-St. Petersburg-Clearwater, Fla.	4.1
Phoenix-Mesa-Scottsdale, Ariz.	4.1
New York-Newark-Jersey City, N.Y.-N.J.-Pa.	4.0
San Antonio-New Braunfels, Texas	4.0
Washington-Arlington-Alexandria, D.C.-Va.-Md.-W.Va.	4.0
Riverside-San Bernardino-Ontario, Calif.	4.0
Philadelphia-Camden-Wilmington, Pa.-N.J.-Del.-Md.	3.9
Baltimore-Columbia-Towson, Md.	3.9
Buffalo-Cheektowaga-Niagara Falls, N.Y.	3.9
Detroit-Warren-Dearborn, Mich.	3.9
Sacramento--Roseville--Arden-Arcade, Calif.	3.9
San Diego-Carlsbad, Calif.	3.9
[illegible]	[illegible]

About 3.9% of San Diegans self-identify as LGBT. If you started asking random San Diegans, how many, on average would you need to ask before someone self-identified as LGBT?

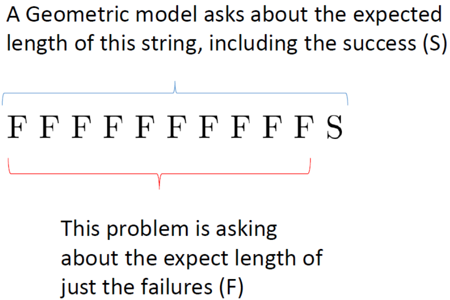
We write $X = Geom(0.039)$.

The problem asks about

$$E(X) = \frac{1}{0.039} \simeq 25.6 \text{ people}$$

Additional Practice

On average, how many non-LGBT people would you encounter when asking like this?



Our answer is just 1 less than the previous problem because it was the expected length when you include the final S, and our problem wants to exclude that: $25.6 - 1 = 24.6$

Additional Practice

What's the probability the first LGBT person you encounter is the sixth person?

We set $X = Geom(p = 0.039)$, and want $P(X = 6)$.

$$P(X = 6) = q^5 \times p = (0.961)^5 \times 0.039 \simeq 0.0320.$$

What's the probability you get your first LGBT person after the second person?
We still have $X = Geom(0.039)$, but now want $P(X > 2)$.

$$\begin{aligned} P(X > 2) &= 1 - P(X = 1) - P(X = 2) \\ &= 1 - 0.039 - 0.961 \times 0.039 \\ &\simeq 0.923. \end{aligned}$$

Additional Practice

In a class of 200, how many people do we expect will LGBT-identify? How much variation can we expect?

Since we have a fixed group size, we use a binomial model:

$$X = Binom(200, 0.039).$$

For expected value, we get

$$E(X) = np = 200 \times 0.039 \simeq 7.8 \text{ LGBT people.}$$

We know

$$SD(X) = \sqrt{npq} = \sqrt{200 \times 0.039 \times 0.961} \simeq 2.7 \text{ LGBT people.}$$

So, it would be common to see counts as low a 7.8 - 2.7 = 5.1, or as high as 7.8 + 2.7 = 10.5 LGBT people.

Additional Practice

In this class of 200, what is the probability that (exactly) three are LGBT?

We set $X = Binom(200, 0.039)$. We want $P(X = 3)$.

$$P(X = 3) = \binom{200}{3}(0.961)^{197}(0.039)^3 \simeq 0.0308.$$

In this class of 200, what is the probability that at least three are LGBT?

The problem wants $P(X \geq 3)$.

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &\simeq 1 - 0.00035 - 0.00284 - 0.01148 \simeq 0.985. \end{aligned}$$

Additional Practice

You work at a business where about 9 customers arrive per hour. What is the probability that no one will arrive in the next 10 minutes?

We are given an average idea in a fixed span (9 customers per hour), so we use the Poisson model.

The challenge here is that the time span in the question (10 minutes) does not match the time span that is given (1 hour). We scale back the given info: 9 per hours equals 9/6 customers per ten minutes.

Setting $X = Poisson(9/6)$, the problem wants $P(X = 0)$.

$$P(X = 0) = \frac{(9/6)^0 \times e^{-9/6}}{0!} \simeq 0.223.$$

Additional Practice

What's the probability that 1 or 2 customers arrive in the next 10 minutes?

We want to find $P(X = 1 \text{ or } X = 2)$.

By disjointness,

$$\begin{aligned} P(X = 1 \text{ or } X = 2) &= P(X = 1) + P(X = 2) \\ &= \frac{(9/6)^1 e^{-9/6}}{1!} + \frac{(9/6)^2 e^{-9/6}}{2!} \\ &\simeq 0.586. \end{aligned}$$