Galvanize statistics short course
Day 2

Hypothesis Testing
Probability review
Marginal, joint, conditional probabilities
Bayes rule
Bayesian vs. Frequentist Statistics
Bayesian Inference

F. Burkholder (credit A. Richards, M. Drury, I. Corneillet)

Welcome back for Day 2

- About me
    - https://www.linkedin.com/in/frankburkholder/
- Course resources:
    - Statistics short course: https://galvanizeopensource.github.io/stats-shortcourse/
    - Mathematics short course: https://galvanizeopensource.github.io/math-essentials-for-data-science/index.html
    - Another statistics presentation: https://github.com/GalvanizeOpenSource/statistics-workshop
    - Galvanize Data Science Prep: https://www.galvanize.com/data-science/prep

Hypothesis Testing - motivating question

FiveThirtyEight.com polled the American public and found that 32% are “very worried” about CoVid-19 infection.

You have a large social network, and you asked 100 of your friends if they are “very worried” about CoVid-19 infection. 26 of them said yes. Are your results statistically different from FiveThirtyEight’s results?*

*Assume that your friends’ demographics are similar to those polled by FiveThirtyEight.

Hypothesis Test

A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

For example:

- My friends' results are the same as FiveThirtyEight's
- My friends' results are not the same as FiveThirtyEight's

In a Hypothesis Test, these two statements are referred to as hypotheses.

Typically, the Null Hypothesis (H0) is the status quo and the alternate (H1) is the more exciting finding.

The Hypotheses must be stated mathematically, and a distribution needs to be associated with the Null Hypothesis.

Hypotheses - motivating example

Null (H0):

- My friends' results are the same as the national average
- $p_{vw,friends} = p_{vw,US} = 0.32$
  (proportion very worried friends = proportion very worried in US)

Alternate (H1):

- My friends' results are not same as the national average
- $p_{vw,friends} \neq p_{vw,US}$
  (proportion very worried friends ≠ proportion very worried in US)

Hypothesis Testing Steps

1. Define your hypotheses.

2. Define your significance level
    - How willing are you to incorrectly reject the Null hypothesis?

3. Compute the probability of your results (or more extreme results) under the Null hypothesis.

4. Compare this probability to your significance level and come to a conclusion.
    - Reject or Fail to Reject the Null Hypothesis

Significance level

The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true.

For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

--Minitab

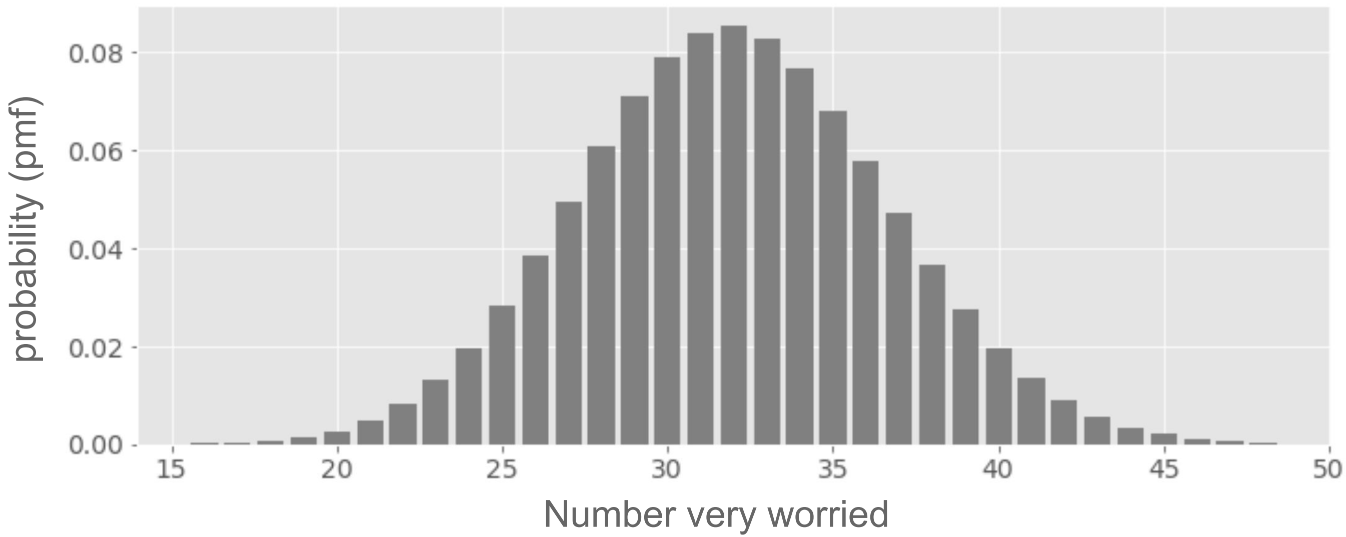
Why a significance level is needed

Here's the null distribution for our motivating example.

It's Binomial with:

n = 100 (sampled 100 people)

p = 0.32 (US average)



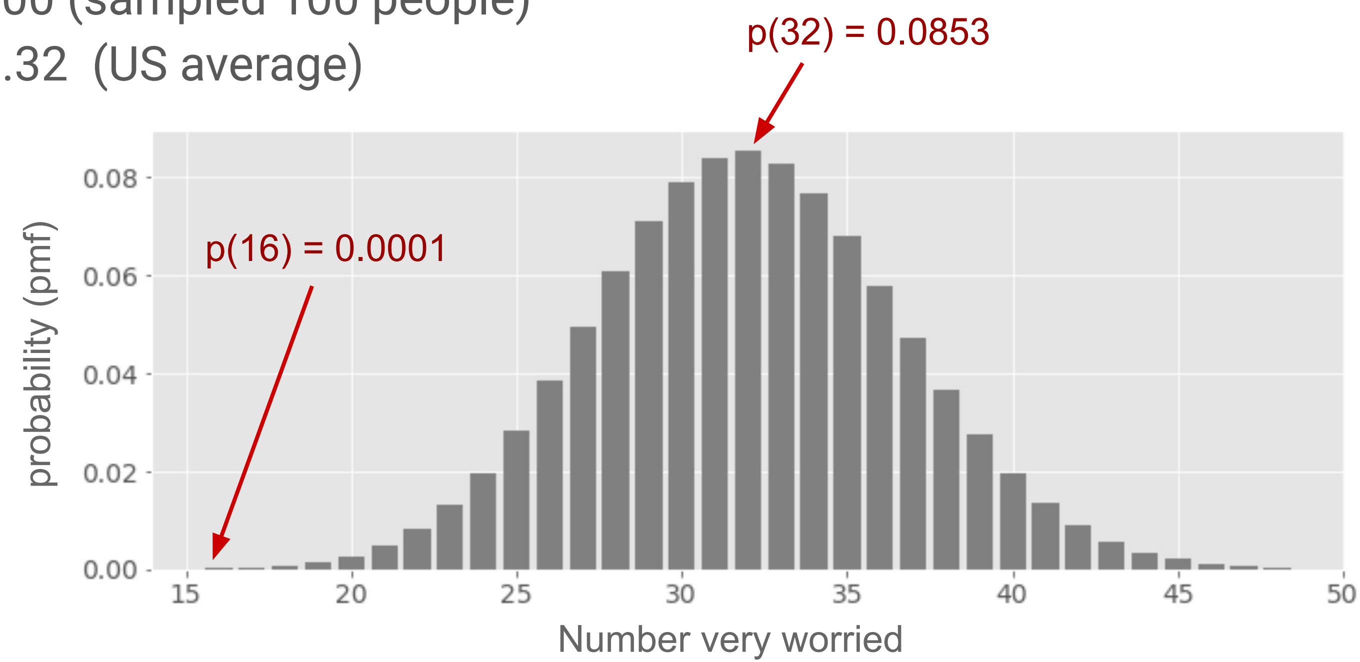
Why a significance level is needed

Here's the null distribution for our motivating example.

It's Binomial with:

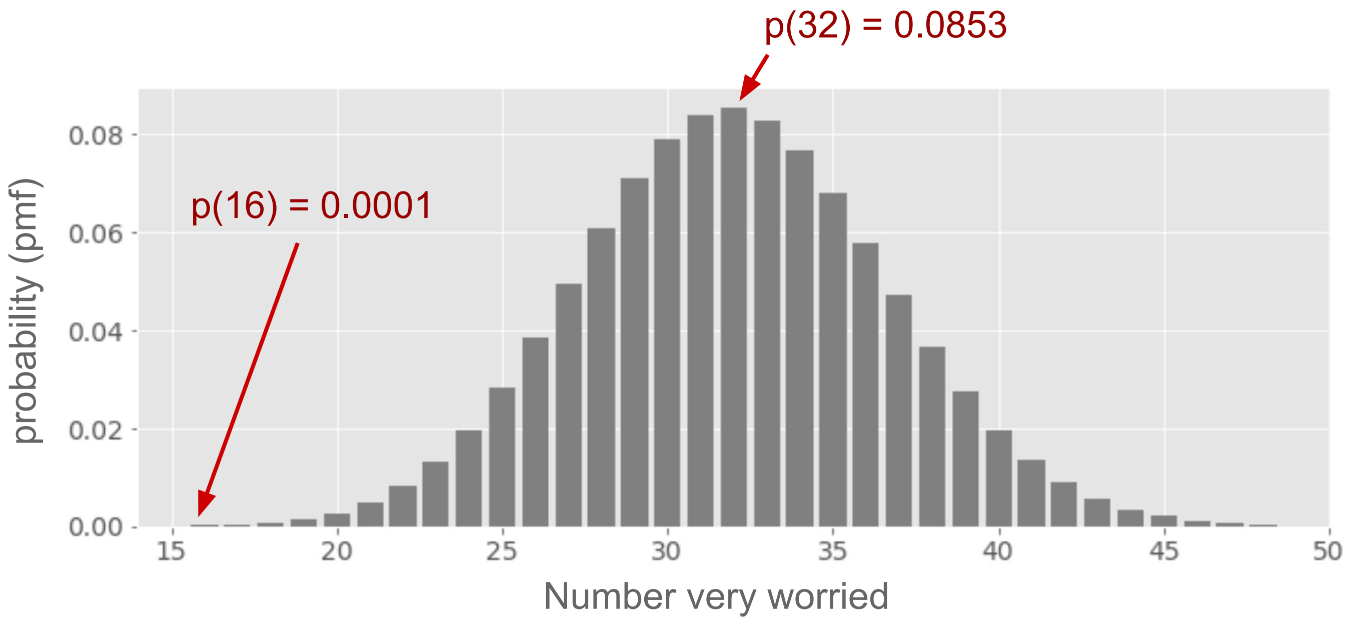
n = 100 (sampled 100 people)

p = 0.32 (US average)



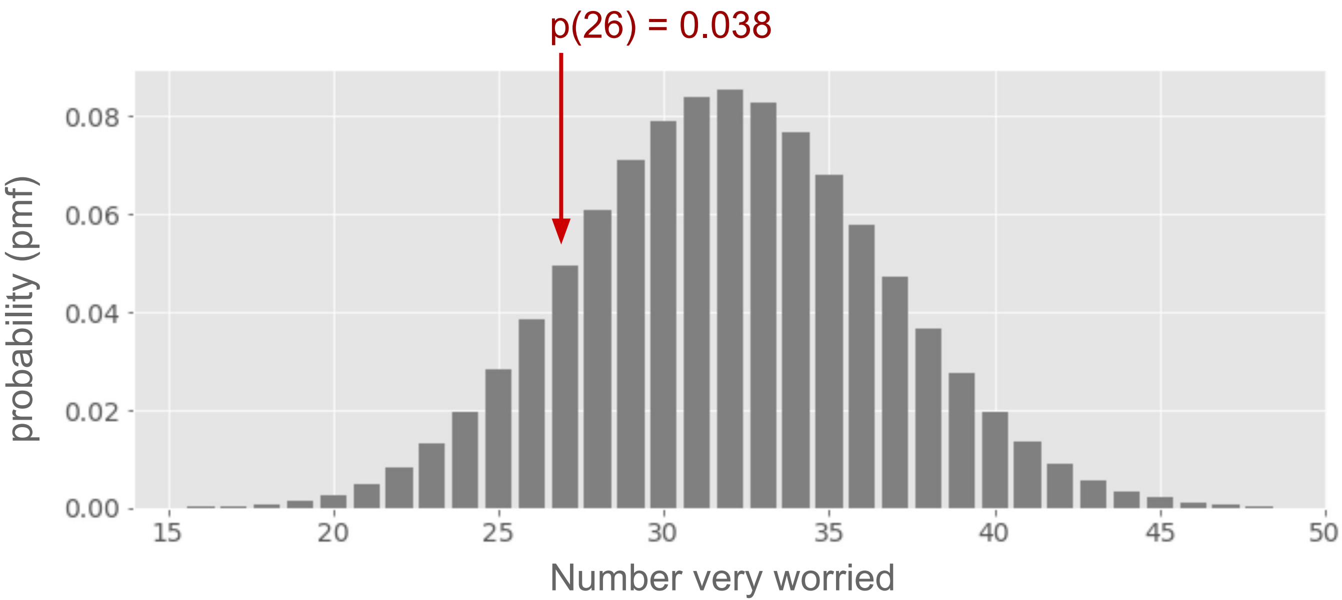
Why a significance level is needed

If we never wanted to incorrectly reject the Null, then we would have to allow for results as extreme as 16 or 48, even though they are very unlikely to come from the Null. So, set a significance level. Typically 5%.



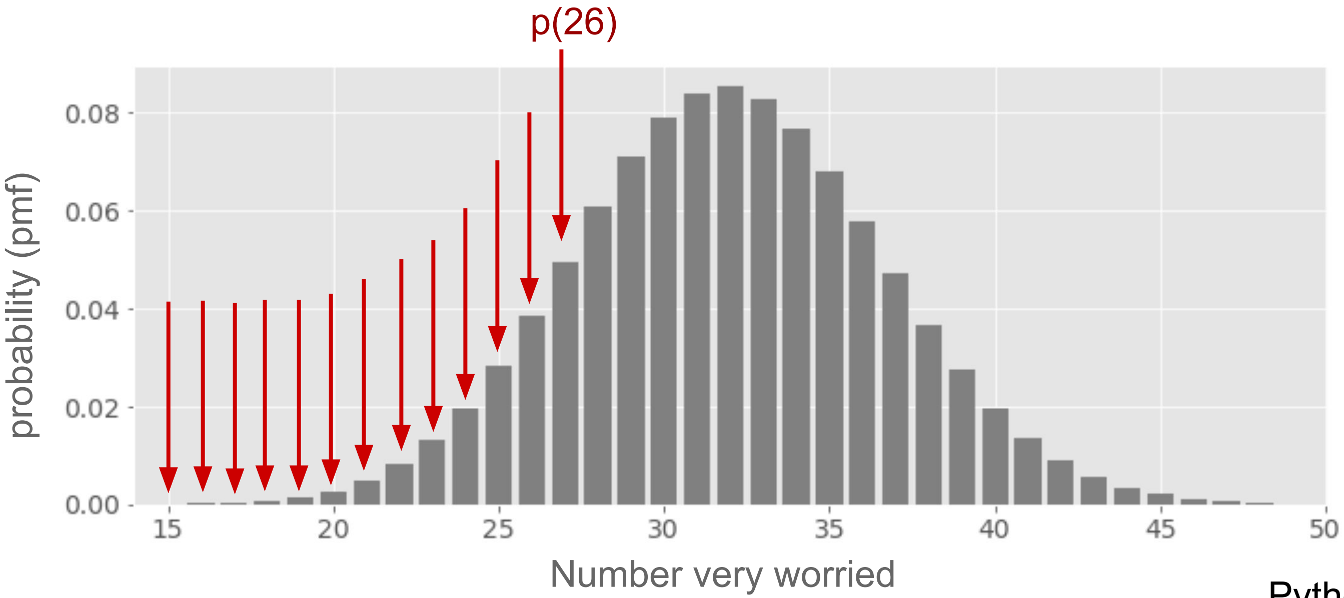
The probability of our results

In our sample of 100, 26 said yes.



The probability of our results (*and more extreme*)

In our sample of 100, 26 said yes. This is less than 32. So is 25, 24, 23, etc.
Need to sum them all up to and including 26.



p(0 - 26) = 0.118

Python

```
n = 100     # we asked 100 of our friends
p = 0.32    # the national average
binomial = stats.binom(n=n, p=p)
print(binomial.cdf(26))
```

Compare probability to significance level, make conclusion

The probability of our results, or more extreme results, given that the Null Hypothesis is true:

p = 0.118 (p-value)

Our significance level, **α = 0.05**

if p > α: Fail to Reject the Null (data likely to have come from Null)

if p <= α: Reject the Null (data unlikely to have come from Null)

Conclusion:

0.118 > 0.05 Fail to Reject Null.

The proportion of our friends that are “very worried” is the same as the US population.

Probability review

What is probability again?

Probability

The study of outcomes in a random process in which the characteristics of the random process are known.

The goal is often understanding what the output of the random process is.

- Expected value
- Minimum, maximum
- Distribution of possible values

For example:

Let’s say we have a fair coin: p(Heads) = 0.5

After 20 flips, what is the probability we get 5 H, 7 H, 10 H, 20 H?

Types of probability

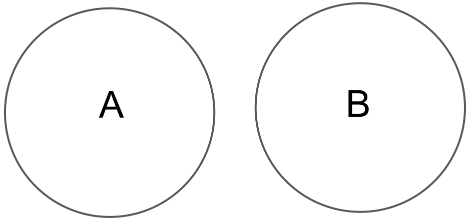
Let's say that there are two events, A and B.

Marginal probability: p(A) or p(B), the probability of an event occurring, not conditional on another event.
Example: what's the probability of drawing a King from a deck of cards?

Joint probability: p(A and B), or $p(A \cap B)$. The probability of event A and event B co-occurring. It's the probability of the intersection of two or more events.
Example: what's the probability of drawing a King (A) and it being of a red suit (B)?

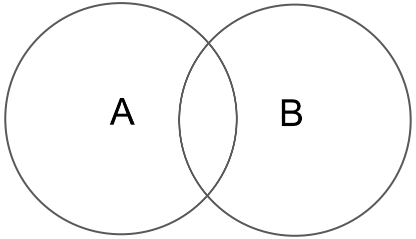
Conditional probability: $p(A|B)$ is the probability of event A occurring, given that event B occurs.
Example: Given a face card is drawn (B), what's the probability it's a King(A)?

Probability illustrated

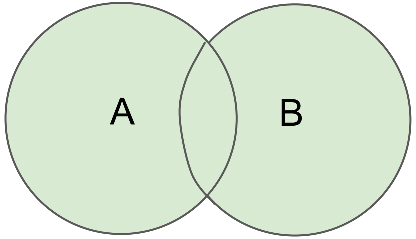


Events A and B (independent).

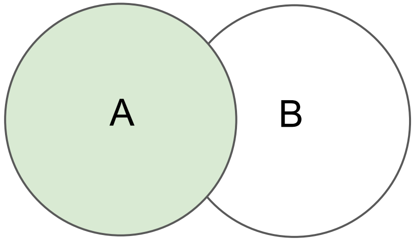
A and B (dependent):



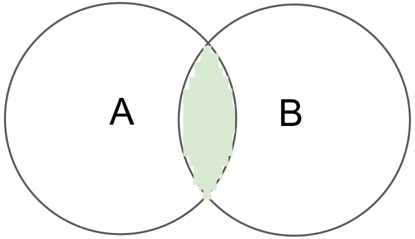
[medium post example]



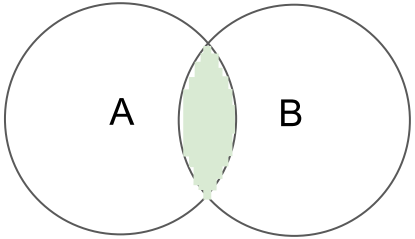
All outcomes (A U B)



marginal probability, p(A): $\frac{A}{A \cup B}$



joint, p(A ∩ B): $\frac{A \cap B}{A \cup B}$



conditional, p(A|B): $\frac{A \cap B}{B}$

Probabilities example

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	
Women	52	48	
(margin)			

Probabilities - Marginal probability

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	100
Women	52	48	100
(margin)	107	93	200

What's the probability that someone prefers Hulu, p(H)?

Probabilities - Marginal probability

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	100
Women	52	48	100
(margin)	107	**93**	**200**

What's the probability that someone prefers Hulu, p(H)?
p(H) = 93/200 = 46.5% (marginal probability)

Probabilities - Joint probability

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	100
Women	52	48	100
(margin)	107	93	200

What's the probability that the next survey respondent is a woman preferring Hulu, $p(W \cap H)$?

Probabilities - Joint probability

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	100
Women	52	**48**	100
(margin)	107	93	**200**

What's the probability that the next survey respondent is a woman preferring Hulu, $p(W \cap H)$? $p(W \cap H) = 48/200 = 24\%$ (joint probability)

Probabilities - Conditional probability

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	100
Women	52	48	100
(margin)	107	93	200

Given that the next survey respondent is a woman, what's the probability that she prefers Hulu, p(H|W)?

Probabilities - Conditional probability

Poll: Do you prefer Netflix or Hulu? Polled 100 men and 100 women.
Results:

	Prefer Netflix	Prefer Hulu	(margin)
Men	55	45	100
Women	52	**48**	**100**
(margin)	107	93	200

Given that the next survey respondent is a woman, what's the probability that she prefers Hulu, $p(H|W) = 48/100 = 48\%$ (conditional probability)

Probabilities: you try

Poll of 400 US residents:
Should the government do more, less, or the same to solve our nation's problems?*

	Same	More	Less
Gen Z	20	70	10
Millennial	20	64	16
Gen X	24	53	23
Boomer	23	49	28

Questions:

1. What's the probability that a respondent thought the US should do more? What type of probability is this?
2. What the probability that a respondent said Less should be done and that they're Gen X? Type of prob?
3. Given that a respondent is Gen Z, what's the probability that they say More is needed? Type of prob?
4. What's the probability of Gen Xers or Boomers saying that Less is needed? Type of prob?
5. What's the probability a respondent is Gen Z or Millennial and say that Less in needed? Type of prob?
6. What's the probability that a respondent says the Same or Less is needed?

* made up data, based on https://www.pewsocialtrends.org/2019/01/17/generation-z-looks-a-lot-like-millennials-on-key-social-and-political-issues/

Conditional probabilities

This is *always* true (conditional or not):

$$Pr(A \cap B) = Pr(A|B) \times Pr(B)$$

This just exercises the statistical chain rule, and it can be extrapolated to more than 2 events.

If events A and B are independent, then:

$$Pr(A|B) = Pr(A)$$

$$Pr(A \cap B) = Pr(A) \times Pr(B)$$

Conditional probabilities

A useful rearrangement:

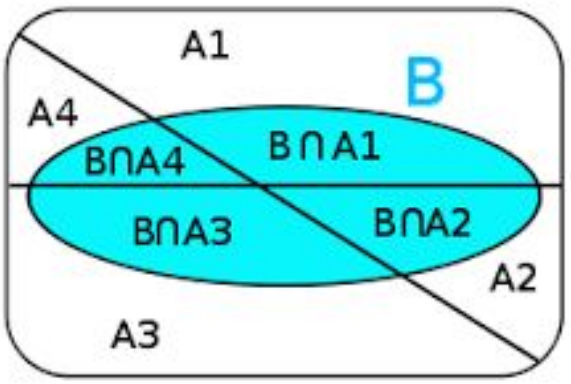
$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Law of total probabilities

$$Pr(B) = \sum_{i=1}^{n} Pr(B \cap A_i)$$

$$Pr(B) = \sum_{i=1}^{n} Pr(B|A_i)Pr(A_i)$$

A



The probability of event B is the sum of the conditional probabilities of event B given a subset of event A for all subsets of event A.

Probabilities example

EXERCISE

Take a moment to think about this question:

- Three types of fair coins are in an urn: HH, HT, and TT

You pick a coin out of the urn and flip it. What's the probability that it's heads?

Probabilities example

EXERCISE

Take a moment to think about this question:

- Three types of fair coins are in an urn: HH, HT, and TT

You pick a coin out of the urn and flip it. What's the probability that it's heads?

Approach 1: Diagram it.

HH

start with coins HH, HT, TT
(equal probability)

HT

TT

Probabilities example

EXERCISE

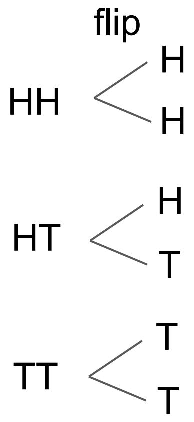
Take a moment to think about this question:

- Three types of fair coins are in an urn: HH, HT, and TT

You pick a coin out of the urn and flip it. What's the probability that it's heads?

Approach 1: Diagram it.

start with coins HH, HT, TT
(equal probability)



Probabilities example

EXERCISE

Take a moment to think about this question:

- Three types of fair coins are in an urn: HH, HT, and TT

You pick a coin out of the urn and flip it. What's the probability that it's heads?

Approach 1: Diagram it.

flip

HH < H, H

HT < H, T

TT < T, T

start with coins HH, HT, TT
(equal probability)

3 of 6 total possible outcomes are H, so 50%.

Probabilities example

EXERCISE

Take a moment to think about this question:

- Three types of fair coins are in an urn: HH, HT, and TT

You pick a coin out of the urn and flip it. What's the probability that it's heads?

Approach 2: Use the law of total probability.

$$Pr(B) = \sum_{i=1}^{n} Pr(B|A_i)Pr(A_i)$$

Probabilities example

EXERCISE

Take a moment to think about this question:

- Three types of fair coins are in an urn: HH, HT, and TT

You pick a coin out of the urn and flip it. What's the probability that it's heads?

Approach 2: Use the law of total probability.

$$
\begin{aligned}
Pr(F_1 = H) &= \sum_{c \in \{HH, HT, TT\}} Pr(F_1 = H | C = c) Pr(C = c) \\
&= 1 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2}
\end{aligned}
$$

Probabilities: you try

- Three types of fair coins are in an urn: HH, HT, and TT
- You pull a coin out of the urn, flip it, and it comes up H
- Q: what is the probability it comes up H if you flip it a second time?

1) Diagram it out to get the right answer, then
2) Use the laws of conditional and total probability to verify it.

Hint: $$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Bayes Rule

From the chain rule:

$$P(A \text{ and } B) = P(A \mid B)P(B)$$

$$P(B \text{ and } A) = P(B \mid A)P(A)$$

Rearranging the law of conditional probability:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Bayes Rule

From the chain rule:

$$P(A \text{ and } B) = P(A \mid B)P(B)$$

$$P(B \text{ and } A) = P(B \mid A)P(A)$$

Rearranging the law of conditional probability:

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

Bayes rule - one of the most important relationships in statistics

Bayes Rule example - are you sick?!

Suppose we have developed a test for a certain disease:

- Only 1% of people have the disease.
- If a person has the disease, the test will be positive 99.9% of the time.
- If a person does not have the disease, the test will be negative 98% of the time.

You get tested for the disease, and the test is positive!

What's the probability that you actually have the disease?

Bayes Rule example - are you sick?!

We are looking for the following conditional probability:

P(have disease | test is positive)

And we know that:

P(have disease) = .01
P(test is positive | have disease) = .999
P(test is positive | don't have disease) = .02 (1 - 0.98)

Bayes Rule example - are you sick?!

Using Bayes rule:

$$P(\text{have disease} \mid \text{test is positive})$$

$$= \frac{P(\text{test is positive} \mid \text{have disease})P(\text{have disease})}{P(\text{test is positive})}$$

We know all the things appearing in the previous formula except P(test is positive).

Bayes Rule example - are you sick?!

We can calculate P(test is positive) using the Law of Total Probability:

$$
\begin{aligned}
&P(\text{test is positive})\\
&= P[\text{test is positive and } \underbrace{(\text{have disease or don't have disease})}_{\text{everything}}]\\
&= P[(\text{test is positive and have disease})\\
&\quad \text{or } (\text{test is positive and don't have disease})]\\
&= P(\text{test is positive and have disease})\\
&\quad + P(\text{test is positive and don't have disease})
\end{aligned}
$$

Bayes Rule example - are you sick?!

We can calculate P(test is positive) using the Law of Total Probability:

$$= P(\text{test is positive} \mid \text{have disease})P(\text{have disease})$$

$$+P(\text{test is positive} \mid \text{don't have disease})P(\text{don't have disease})$$

$$\text{(apply conditional probability formula twice.)}$$

$$= \; .999 \times .01 + \; .02 \times .99$$

Bayes Rule example - are you sick?!

Now that we know P(test is positive) we can solve the problem:

$$P(\text{have disease} \mid \text{test is positive})$$
$$= \frac{P(\text{test is positive} \mid \text{have disease})P(\text{have disease})}{P(\text{test is positive})}$$
$$= \frac{.999 \times .01}{.999 \times .01 + .02 \times .99}$$
$$= .34$$
$$= 34\%$$

Bayes Rule example - are you sick?!

The probability we have the disease is only 34%, even though we received a positive test.

This kind of result is unintuitive to most people, a mental bias called the **base rate fallacy.**

Pretty much everyone's intuition says that it should be much more likely that the person does have the disease after a test comes back positive.

Most undervalue the prior information that P(have disease) = .01
It takes a lot of evidence to make an unlikely situation likely.

Bayes Rule example - understanding the terms

P(have disease) is called the **prior probability**.
It's what we know before collecting evidence/data.

P(test is positive | have disease) is called the **likelihood**.
It's the strength of the evidence/data we collected.

P(have disease | test is positive) is called the **posterior**.
It's what we know, after collecting evidence/data.

$$\underbrace{P(A \mid B)}_{\text{posterior}} = \frac{\overbrace{P(B \mid A)}^{\text{likelihood}}\overbrace{P(A)}^{\text{prior}}}{P(B)}$$

Bayes Rule: you try

A Case of Two Cookie Jars

Let’s say you have two cookie jars in front of you. Both cookie jars contain vanilla and chocolate cookies, but in different proportions. The first jar contains 30 vanilla cookies and 10 chocolate cookies. The second jar contains 20 vanilla cookies and 20 chocolate cookies.

You’re blindfolded and told to pick a cookie at random from one of the two jars. After picking, you’re told that you selected a vanilla cookie.

What is the probability that you picked the cookie from either jar?

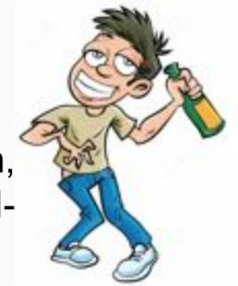
source

Frequentist vs. Bayesian paradigms

Experiment 1:

A fine classical musician says he’s able to distinguish Haydn from Mozart.
Small excerpts are selected at random and played for the musician.
Musician makes 10 correct guesses in exactly 10 trials.

Experiment 2:



Drunken man says he is psychic and can correctly guess what face of the coin will fall down, mid-air. Coins are tossed and the drunken man shouts out guesses while the coins are mid-air. Drunken man correctly guesses the outcomes of the 10 throws.

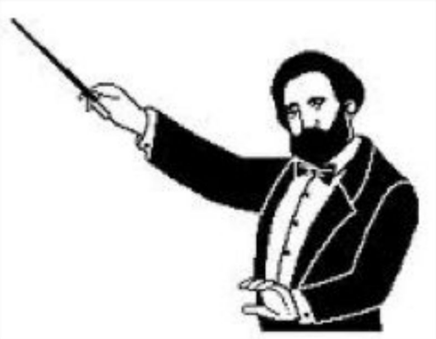
Adapted example from Jim Berger’s book, The Likelihood Principle.

Frequentist vs. Bayesian paradigms

Frequentist: probability can be defined as the long-run frequency of an outcome

Bayesian: a measure of a degree of belief

Frequentist vs. Bayesian paradigms



Frequentist: “They’re both so skilled! I have **as much confidence** in musician’s ability to distinguish Haydn and Mozart as I do the drunk’s psychic ability to predict coin tosses”

Bayesian: “I’m not convinced by the drunken man…”

The Bayesian approach is to incorporate prior knowledge into the experimental results.

Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Rule

posterior probability

likelihood

prior probability

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

marginal likelihood

Bayesian Approach

$$P(\text{psychic}|\text{correct}) = \frac{P(\text{correct}|\text{psychic})P(\text{psychic})}{P(\text{correct})}$$

$$= \frac{1.0 * 0.0001}{0.5^{10}}$$

Very subjective!

$$= 10.2\%$$

Bayesian Approach to Parameter Estimation

Let’s say you’re trying to estimate the CTR (Click-Through-Rate) of a web page your company has designed. When a user comes to the web page, they either click on a desired link (giving you a 1 in your dataset), or they don’t (giving you a 0 in your dataset).

Enabling Bayes for proportions: Beta distribution

The Beta Distribution: $\mathrm{Beta}(\alpha, \beta)$

The beta distribution has two hyper-parameters (also known as "shape parameters"):

- $\alpha > 0$: we will use this to encode the number of "successes" of a website (more on that later)
- $\beta > 0$: we will use this to encode the number of "failures" of a website (more on that later)

Support: $x \in [0, 1]$

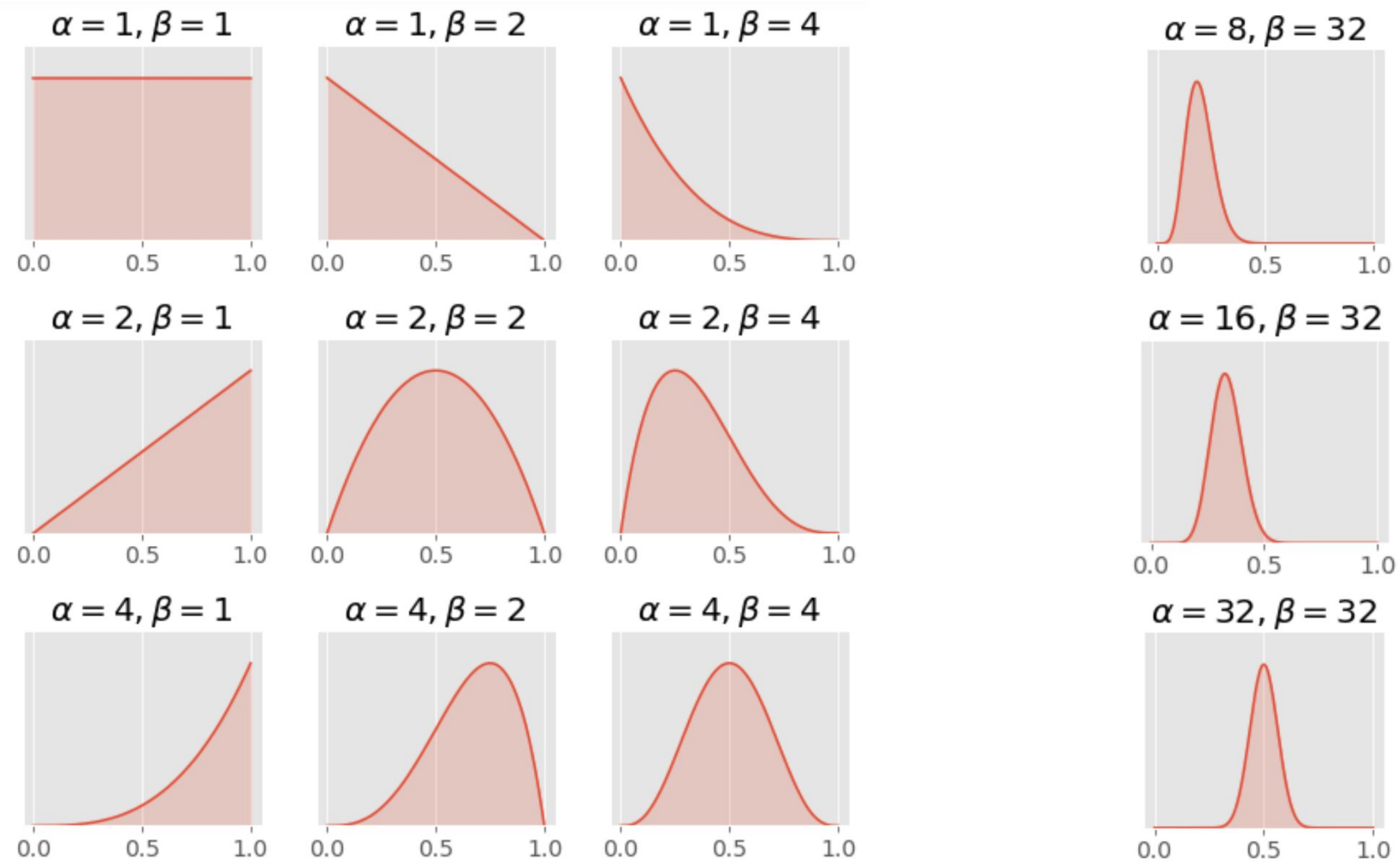
PDF: $f(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$

where $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

where $\Gamma(x)$ is an extension of the factorial function which works for a wider range of input (like any positive real value, which is why it works here).

Beta distribution on Wikipedia

What the Beta looks like for (α, β)



Using Bayes Rule with Binomial & Beta

Beta: models CTR from Prior & Posterior.

$\alpha = 1 + \text{number of conversions on our website}$
$\beta = 1 + \text{number of misses on our website}$

Binomial: models likelihood of our stream of clicks: (0s and 1s)

Bayes rule:

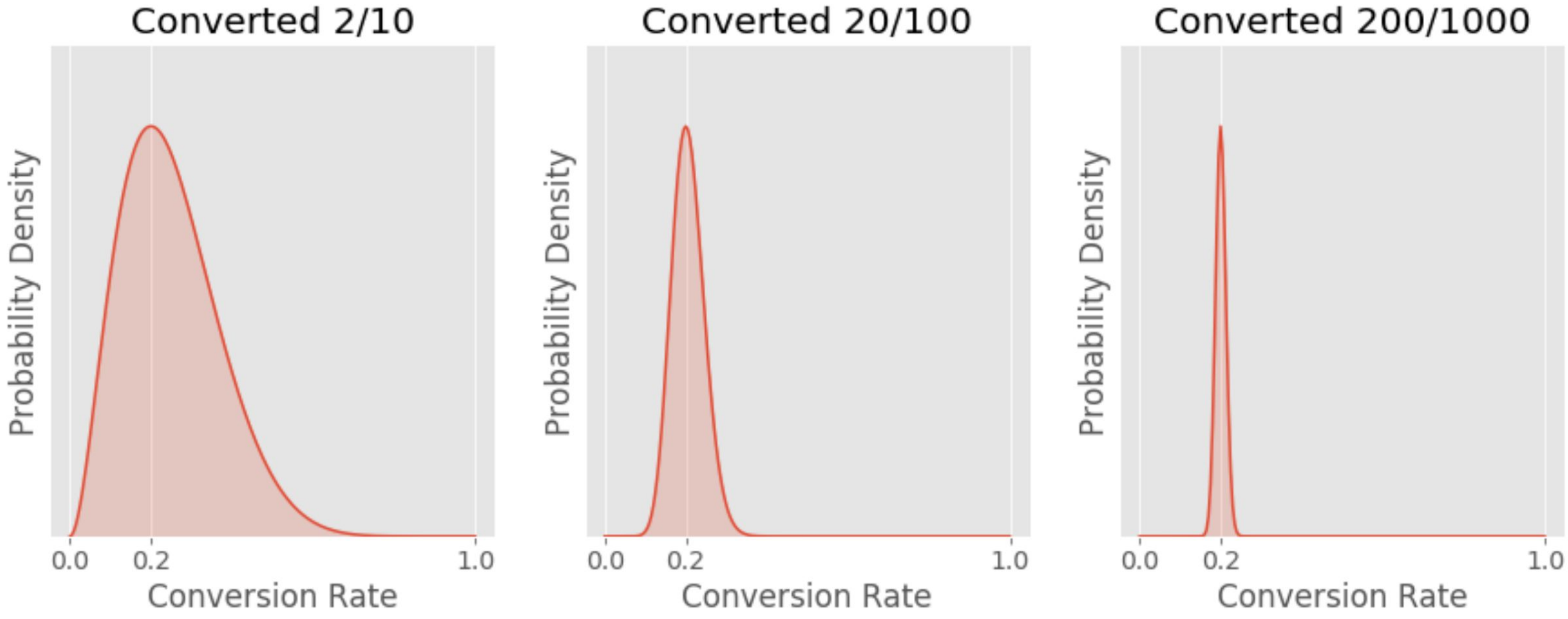
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$ $\longleftarrow$ Normalizing constant

$$P(CTR \mid clicks) \propto P(clicks \mid CTR) * P(CTR)$$

$$\text{Beta} \propto \text{Binomial} * \text{Beta}$$

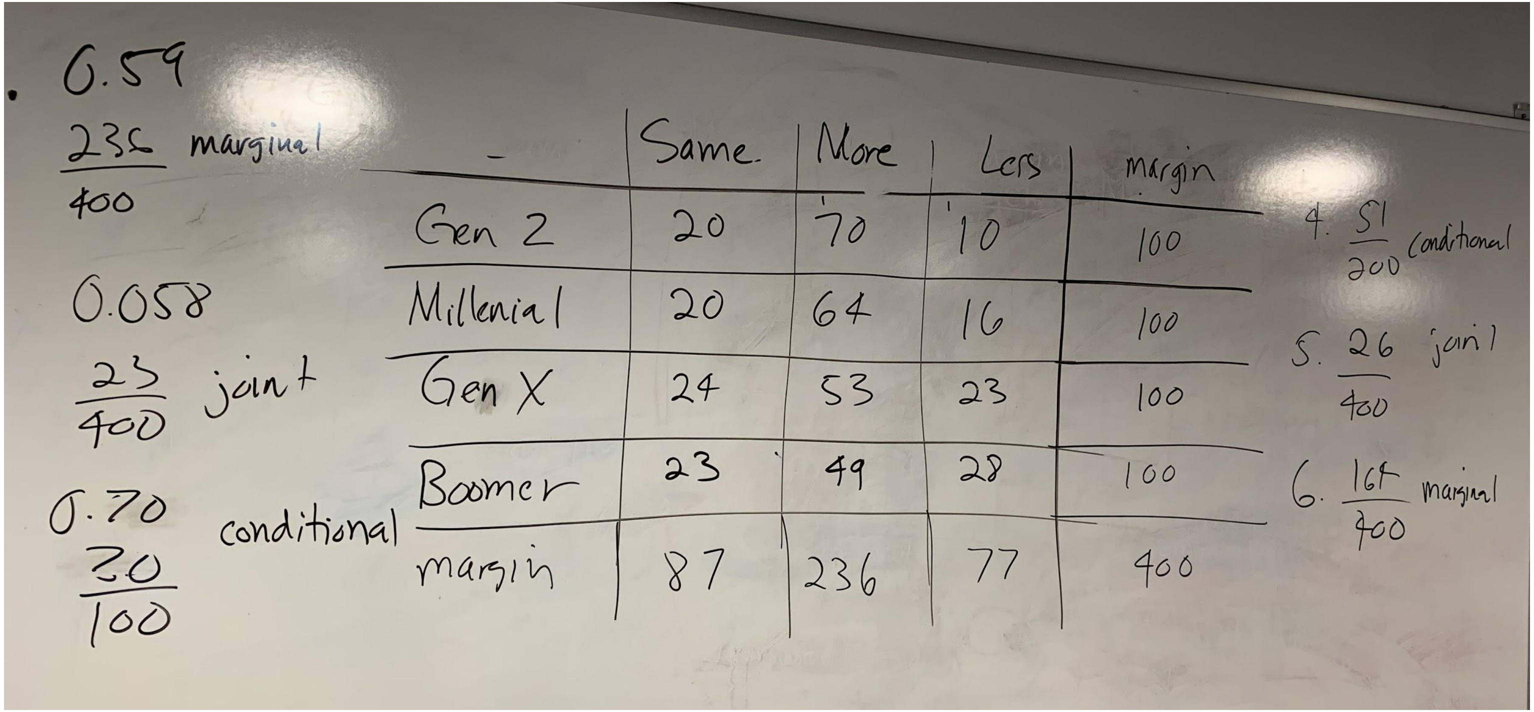
This works because Beta is a conjugate prior of the Binomial distribution.

Application of Bayesian Inference



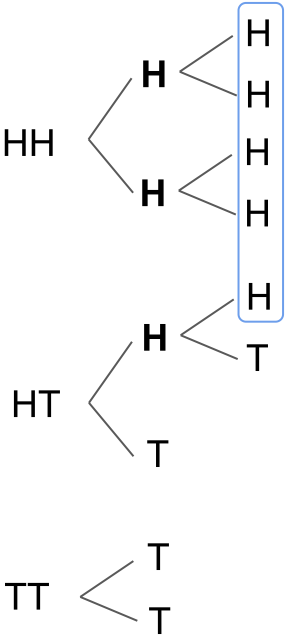
Thank you!

Probabilities table answer



Probabilities example answer

- Three types of fair coins are in an urn: HH, HT, and TT
- You pull a coin out of the urn, flip it, and it comes up H
- Q: what is the probability it comes up H if you flip it a second time?



5/6

1. $Pr(F_1 = H \cap F_2 = H)$
2. $Pr(F_1 = H)$
3. $Pr(F_2 = H | F_1 = H) = \frac{Pr(F_1=H, F_2=H)}{Pr(F_1=H)}$

as

$$Pr(F_1 = H \cap F_2 = H) = \sum_{c \in \{HH, HT, TT\}} Pr(F_1 = H \cap F_2 = H | C = c) Pr(C = c)$$
$$= 1 \times \frac{1}{3} + \frac{1}{4} \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{5}{12}$$

$$Pr(F_1 = H) = \sum_{c \in \{HH, HT, TT\}} Pr(F_1 = H | C = c) Pr(C = c)$$
$$= 1 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2}$$

$$\Pr(F_2 = H | F_1 = H) = \frac{5/12}{1/2} = \frac{5}{6}$$