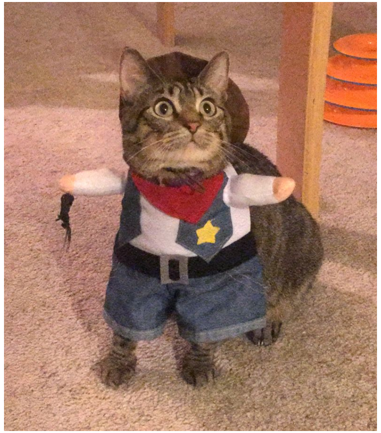
INFO 2950: Intro to Data Science

Lecture 19
2023-11-01

Agenda



1. Joint and marginal probabilities
2. Hypothesis Tests for Independence
3. Conditional probabilities
4. Bayes Theorem

Interview question alert!

@Allison inspired by u I've been asking this as an introductory interview question

😆 2

Everyone ive asked so far has taken at least like 15 secs to answer and done it by brute force

😮

Interview question alert!

What is the probability that you roll two fair dice and the sum of their faces is 7?

@Allison inspired by u I've been asking this as an introductory interview question

😆 2

Everyone ive asked so far has taken at least like 15 secs to answer and done it by brute force

😮

Interview question alert!

Brute force method:

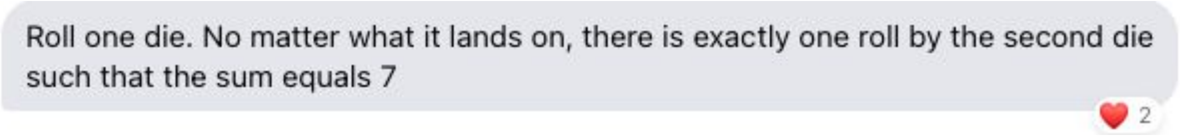
- [1,6],[2,5],[3,4],[4,3],[5,2],[6,1] → 6 ways to roll a 7
- 6*6 = 36 total ways to roll two dice
- 6/36 = ⅙ probability of rolling a 7

Interview question alert!

Brute force method:

- [1,6],[2,5],[3,4],[4,3],[5,2],[6,1] → 6 ways to roll a 7
- 6*6 = 36 total ways to roll two dice
- 6/36 = ⅙ probability of rolling a 7

Non-brute-force method:


Roll one die. No matter what it lands on, there is exactly one roll by the second die such that the sum equals 7

❤️ 2

Probabilities on variable(s)

- So far, we've focused on probabilities on a single variable
  - Probability that [you roll a 7 from two dice]
  - $P(X=k)$

Probabilities on variable(s)

- So far, we've focused on probabilities on a single variable
  - Probability that [you roll a 7 from two dice]
  - P(X=k)

- What if we care about probabilities on multiple variables?
  - Probability that [you roll a 7 from two dice] **and/or/given** [you roll a 1 first]

Probabilities and notation

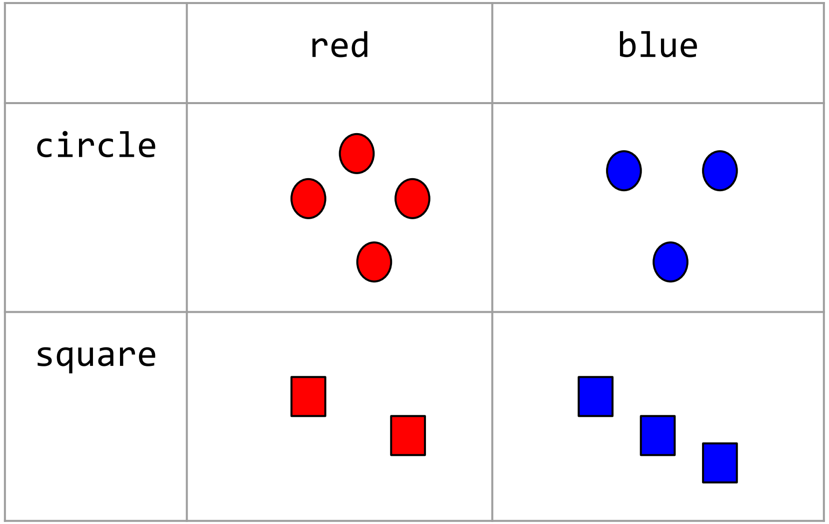
Joint probability of two or more variables

$P(X, Y)$

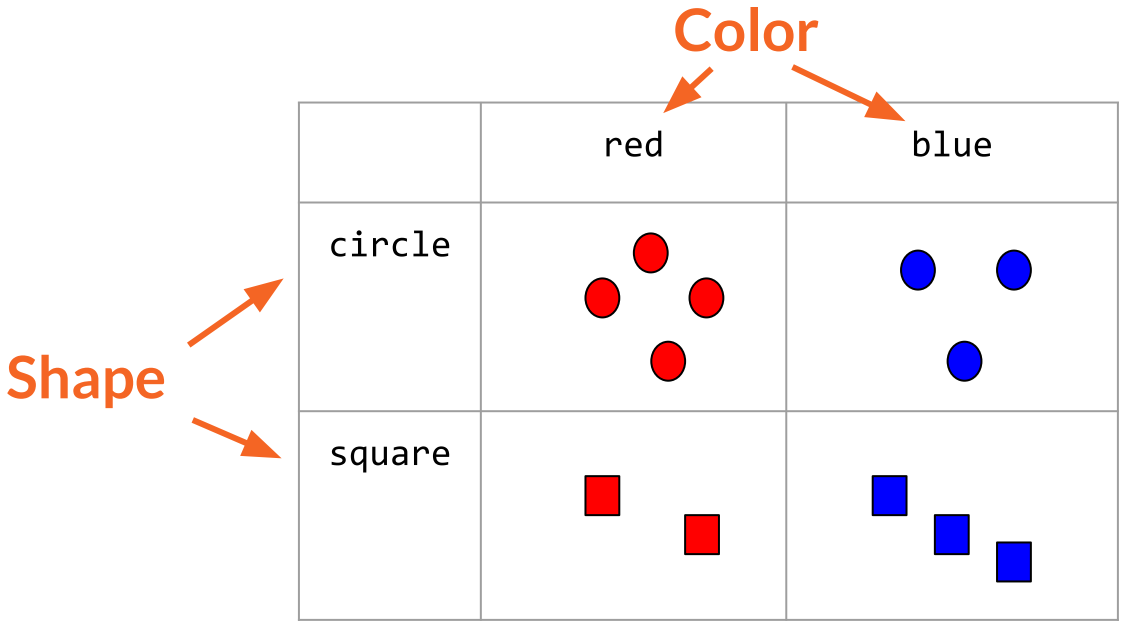
Conditional probability of one variable *given* a value of another variable

$P(X \mid Y=y)$

From counting things to probability

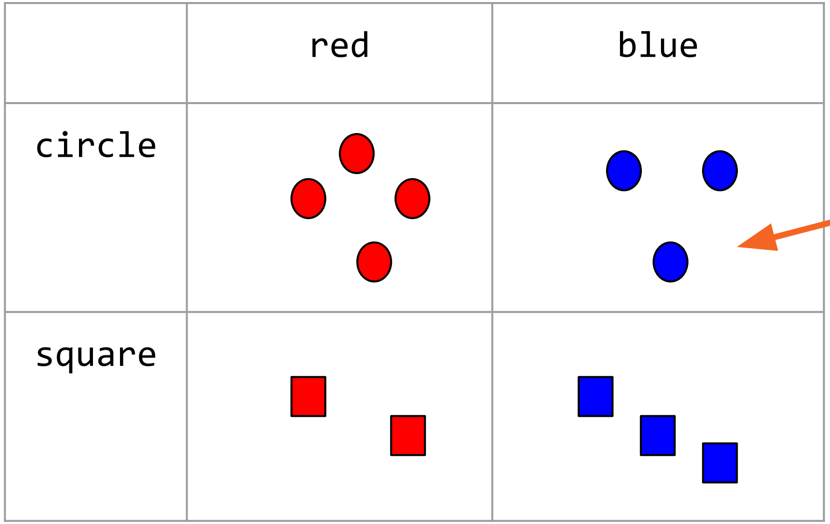


Contingency tables



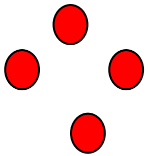
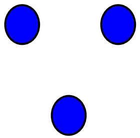
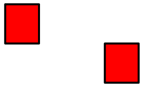
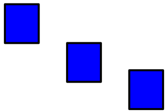
Each cell represents a combination of two variables

Contingency tables



Value is an integer counting the frequency of a combination

Normalize to joint probability P(Shape, Color)

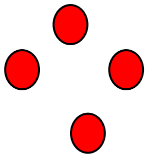
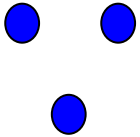
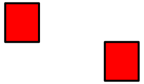
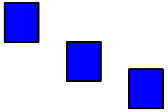
	red	blue
circle		
square		

→

	red	blue
circle	?	?
square	?	?

"Normalize" = Divide everything by the total sum

Normalize to joint probability P(Shape, Color)

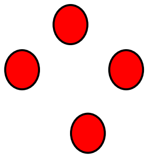
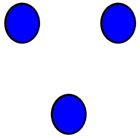
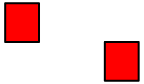
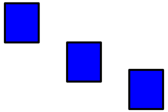
	red	blue
circle		
square		

→

	red	blue
circle	?	?
square	?	?

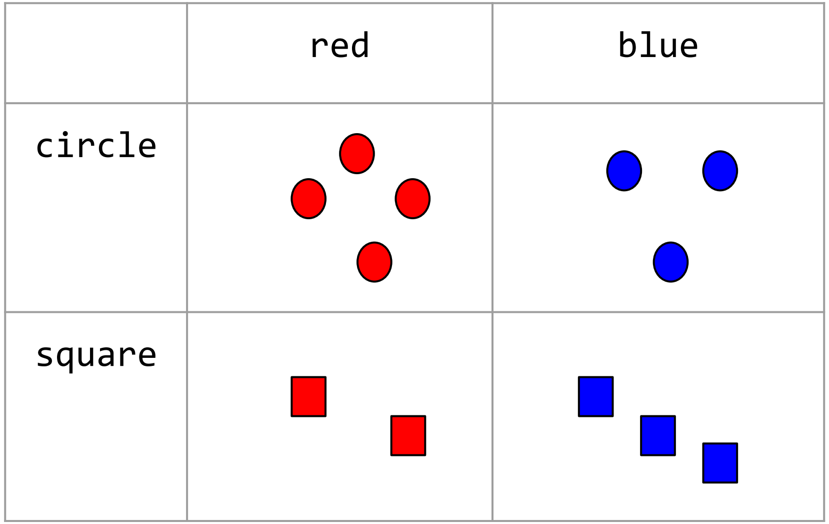
Hint: the joint probability denominator is the total # things in the matrix (4+3+2+3 = ?)

Normalize to joint probability P(Shape, Color)

	red	blue
circle		
square		

	red	blue
circle	1/3	1/4
square	1/6	1/4

Normalize to joint probability P(Shape, Color)



→

	red	blue
circle	1/3	1/4
square	1/6	1/4

There are 4 red circles, divided by 12 → ⅓
There are 3 blue squares, divided by 12 → ¼
etc.

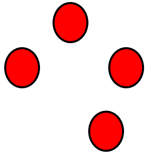
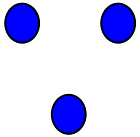
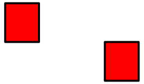
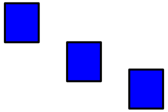
Can we reverse normalization?

	red	blue
circle	?	?
square	?	?

←

	red	blue
circle	1/3	1/4
square	1/6	1/4

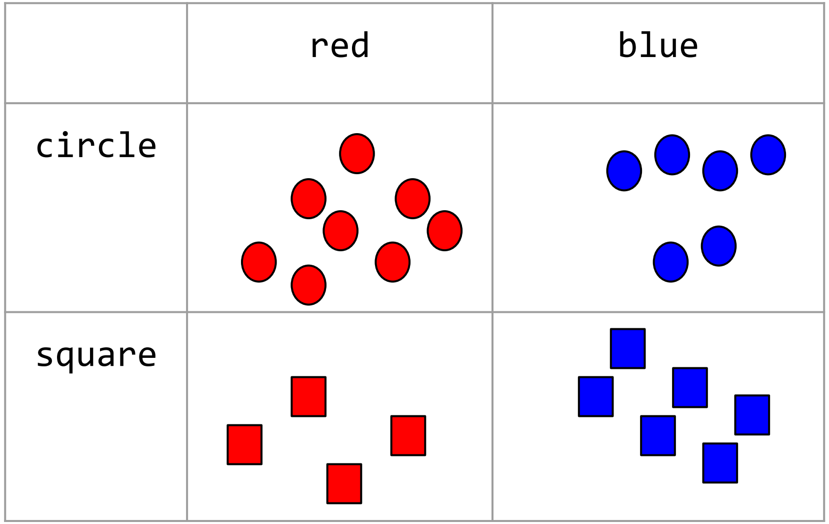
Normalization loses information

	red	blue
circle		
square		

	red	blue
circle	1/3	1/4
square	1/6	1/4

This is one possible assortment…

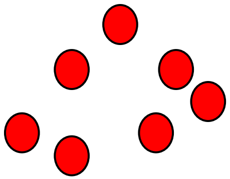
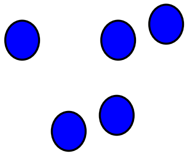
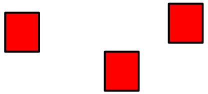
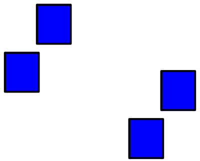
Normalization loses information



	red	blue
circle	1/3	1/4
square	1/6	1/4

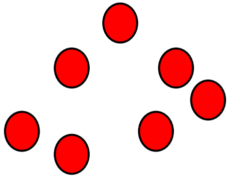
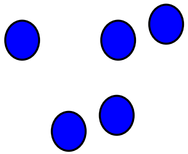
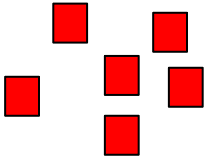
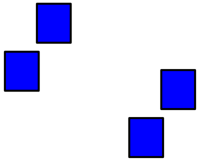
But here's another possibility!

Counting degrees of freedom

	red	blue
circle		
square		

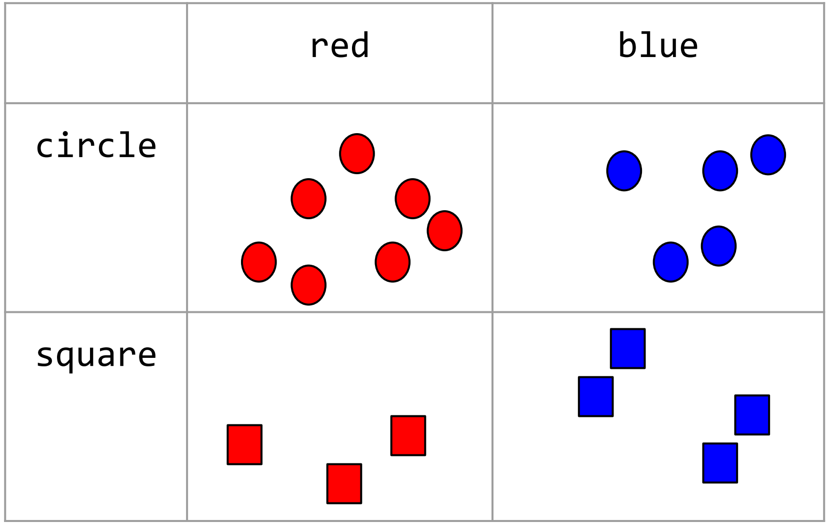
In a 2x2 contingency table, without any other constraints, you can put any four integers into the cells

Counting degrees of freedom

	red	blue
circle		
square		

In a 2x2 contingency table, without any other constraints, you can put any four integers into the cells

Counting degrees of freedom



In a 2x2 contingency table, without any other constraints, you can put any four integers into the cells

	red	blue
circle	1/3	1/4
square	1/6	?

In a joint probability table, if you know three of the fractions, the fourth is **determined**

Counting degrees of freedom

	red	blue
circle	1/3	1/4
square	1/6	?

What is this value?

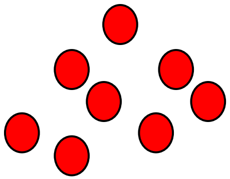
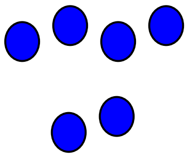
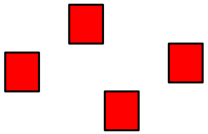
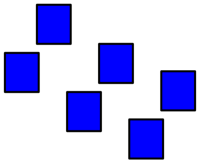
Counting degrees of freedom (*df*)

The combinations of colors and shapes are mutually exclusive. If something is NOT [red circle / blue circle / red square], then it must be a blue square.

$$1 - (\tfrac{1}{3} + \tfrac{1}{6} + \tfrac{1}{4}) = 1 - \tfrac{3}{4} = \tfrac{1}{4}$$

	red	blue
circle	1/3	1/4
square	1/6	1/4

Normalization loses information

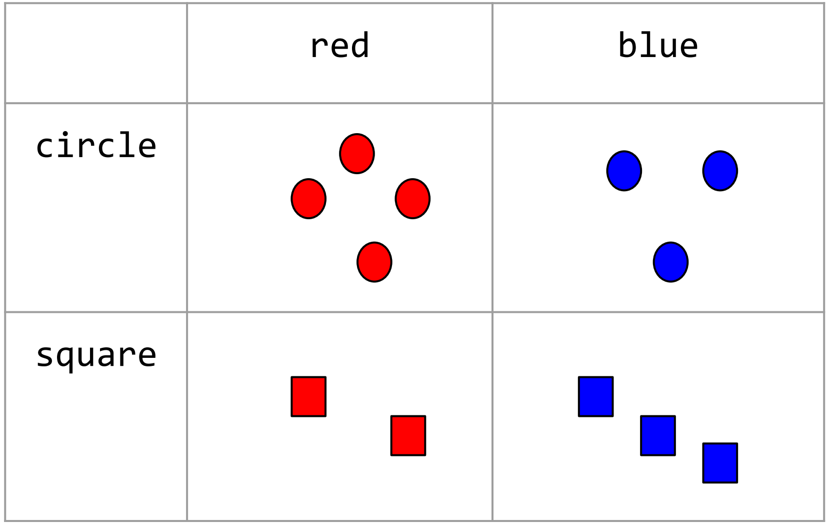
	red	blue
circle		
square		

	red	blue
circle	1/3	1/4
square	1/6	1/4

Four free parameters for counts (df=4)

Three free parameters (df=3) for joint probability table

Reminder: joint probability is P(Shape, Color)

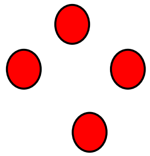
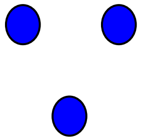
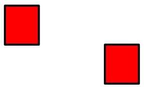
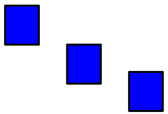


	red	blue
circle	1/3	1/4
square	1/6	1/4

Marginal probabilities sum over one axis

Axis 1

Axis 0

P(S,C)	red	blue
circle		
square		

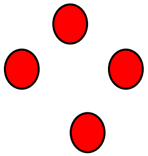
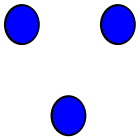
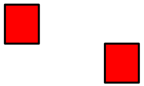
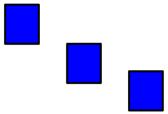
P(S)	
circle	?
square	?

P(C)	red	blue
	?	?

Marginalizing over shapes (counting without joint distributions)

Axis 1

Axis 0

P(S,C)	red	blue
circle		
square		

P(C)	red	blue
	(4+2)/12 = ½	(3+3)/12 = ½

Marginalizing over shapes (summing the joint distributions of shapes)

Axis 1 →

Axis 0 ↓

	red	blue
circle	1/3	1/4
square	1/6	1/4

P(S)	
circle	?
square	?

P(C)	red	blue
	⅓ + ⅙	¼ + ¼

Marginalizing over shapes

Axis 1

Axis 0

	red	blue
circle	1/3	1/4
square	1/6	1/4

P(S)	
circle	?
square	?

P(C)	red	blue
	½	½

Summing joint distributions gives us the correct probability!

Marginal probabilities sum over one axis

Axis 1 →

Axis 0 ↓

	red	blue
circle	1/3	1/4
square	1/6	1/4

P(S)	
circle	?
square	?

P(C)	red	blue
	½	½

Marginalizing over colors

Axis 1

Axis 0

	red	blue
circle	1/3	1/4
square	1/6	1/4

P(S)	
circle	?
square	?

P(C)	red	blue
	½	½

Marginalizing over colors

Axis 1 →

Axis 0 ↓

	red	blue
circle	1/3	1/4
square	1/6	1/4

P(S)	
circle	7/12
square	5/12

P(C)	red	blue
	1/2	1/2

Marginal probabilities lose degrees of freedom

Axis 1 →

Axis 0 ↓

Each marginal distribution has **one** free parameter (df=1), so together they have **two** parameters

P(S)	
circle	7/12
square	5/12

P(C)	red	blue
	1/2	1/2

Can we reverse marginalization?

	red	blue
circle	?	?
square	?	?

P(S)	
circle	7/12
square	5/12

P(C)	red	blue
	6/12	6/12

Can we reverse marginalization?

	red	blue
circle	P(red)* P(circle)	P(blue)* P(circle)
square	P(red)* P(square)	P(blue)* P(square)

P(S)	
circle	7/12
square	5/12

P(C)	red	blue
	6/12	6/12

Can we reverse marginalization?

	red	blue
circle	42/144	42/144
square	30/144	30/144

P(S)	
circle	7/12
square	5/12

P(C)	red	blue
	6/12	6/12

Marginalization is not (always) reversible

	red	blue
circle	1/3	1/4
square	1/6	1/4

$\neq$

	red	blue
circle	42/144	42/144
square	30/144	30/144

Marginalization is not (always) reversible

	red	blue
circle	1/3	1/4
square	1/6	1/4

$\neq$

	red	blue
circle	42/144	42/144
square	30/144	30/144

P(shape, color) ≠ P(shape) P(color)

Marginalization is not (always) reversible

	red	blue
circle	1/3	1/4
square	1/6	1/4

≠

	red	blue
circle	42/144	42/144
square	30/144	30/144

The joint probability is only equal to the product of the marginal probabilities if the variables are **independent!**

https://www.facebook.com/statmemes/

Marginalization is not (always) reversible

If the joint probability is not equal to the product of the marginals, does that imply that the variables **cannot** be independent?

Marginalization is not (always) reversible

If the joint probability is not equal to the product of the marginals, does that imply that the variables **cannot** be independent?

Not necessarily! Just like flipping five heads in a row doesn't mean the coin is biased, we expect some degree of random variation in a contingency table.

The critical question is how much evidence the deviation from the expected counts provides that something else is going on.

Example: flipping two coins

	heads	tails
heads	0	6
tails	2	2

I flip two coins 10 times and count the number of pairs

Null hypothesis: P(heads, heads) = P(heads)P(heads)

Alternative hypothesis: P(heads, heads) ≠ P(heads)P(heads)

Are the coins independent?

DeGraffenreid v. General Motors (1976)

- GM did not hire Black women before 1964
- In the early 1970s recession, GM did layoffs by seniority → all the Black women were laid off
- 5 Black women sued GM over discrimination by gender *and* race
- Unsuccessful because the court didn't know how to deal with the *intersection* of gender and race

https://www.vox.com/the-highlight/2019/5/20/18542843/intersectionality-conservatism-law-race-gender-discrimination

DeGraffenreid v. General Motors (1976)

> "The legislative history surrounding Title VII does not indicate that the goal of the statute was to create a new classification of 'black women' who would have greater standing than, for example, a black male. The prospect of the creation of new classes of protected minorities, governed only by the mathematical principles of permutation and combination, clearly raises the prospect of opening the hackneyed Pandora's box."

- Judge Harris Wangelin's ruling against the plaintiffs

https://www.vox.com/the-highlight/2019/5/20/18542843/intersectionality-conservatism-law-race-gender-discrimination

Hypothesis testing for intersectionality

GM argued that they were not discriminating against black people because they hired black men, and they were not discriminating against women, because they hired white women.

Null hypothesis: **P(Black, woman)** = **P(Black)P(woman)**

Alternative hypothesis: **P(Black, woman)** < **P(Black)P(woman)**

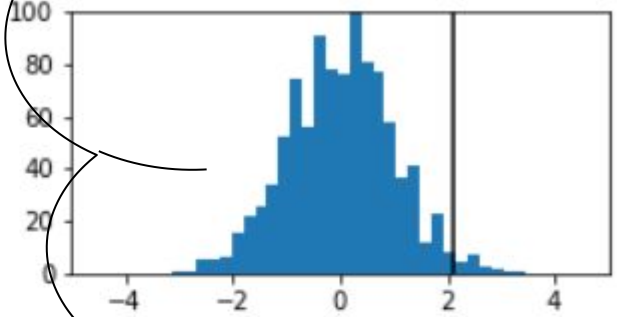
*Sloppy notational shorthand: everything in parentheses refers to the probability of being hired given the worker is Black / female)

Independence

- It's often useful to know "in the real world" if things are independent
  - Joint probabilities are easier if A and B are independent: $P(A,B) = P(A) * P(B)$

- How do we test for independence?

Hypothesis Testing



$$Z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Null (boring) hypothesis H_0
- Alternative (spooky) hypothesis H_1
- Plot where your value is relative to the distribution
- Decide whether you reject the null
  (Don't assume anything with $p < 0.05$ is always significant!)

Hypothesis Testing: Independence

- **Null (boring) hypothesis: A and B are independent**
  - The probability of A occurring is the same for different values of B

- **Alternative (spooky) hypothesis: A and B are not independent**
  - The probability of A occurring depends on B

Testing for Independence

- Use a **contingency table** in order to run a **chi-squared test**
  - A contingency table allows you to look at the frequency distribution of different variables
  - A chi(χ)-squared test is a hypothesis test used when the null is χ^2 distributed (the sum of squares of normal random variables)

Why χ^2 distributed?

- Our hypothesis test's *test statistic* will be a sum of a squared value. Instead of a "z-score" we now have:

$$\sum_{(i \cdot j)} \frac{(O - E)^2}{E}$$

- O = observed values
- E = expected values
- i = the number of rows in the table
- j = the number of columns in the table

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

These are Observed values (O)

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

How do we get Expected (E) values from Observed (O) values?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Let’s calculate the expected frequency of CC students volunteering 1-3 hours.

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

Let's calculate the expected frequency of CC students volunteering 1-3 hours.

If these are independent, then P(CC and 1-3hr) = P(CC) * P(1-3hr)

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Let's calculate the expected frequency of CC students volunteering 1-3 hours.

If these are independent, then P(CC and 1-3hr)
= P(CC) * P(1-3hr)
= [255/839] * P(1-3hr)

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Let's calculate the expected frequency of CC students volunteering 1-3 hours.

If these are independent, then P(CC and 1-3hr)
= P(CC) * P(1-3hr)
= [255/839] * [298/839]

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	$\boxed{839}$

P(CC and 1-3hr) = E/839 = [255/839] * [298/839]

→ E = [255] * [298] / [839]

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

P(CC and 1-3hr) = E/839 = [255/839] * [298/839]

→ **E** = [255] * [298] / [839] = [row total] * [col total] / [total] = 90.57

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

Observed value O:

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

In contrast to O=111, we calculated an Expected value E for College Students volunteering 1-3 hours
= [255] * [298] / [839]
= [row total] * [col total] / [total] = 90.57

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Contingency table math

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

Observed values (O)

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours
Community College Students	90.57	115.19	49.24
Four-Year College Students		131.00	56.00
Nonstudents	104.42	132.81	56.77

Expected values (E)
(calculate assuming student type and # hrs are independent)

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Hint: row * col / total

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

Observed values (O)

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours
Community College Students	90.57	115.19	49.24
Four-Year College Students	____	131.00	56.00
Nonstudents	104.42	132.81	56.77

Expected values (E)
(calculate assuming student type and # hrs are independent)

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

$290 * 298 / 839 = 103.00$

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

Observed values (O)

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours
Community College Students	90.57	115.19	49.24
Four-Year College Students	103.00	131.00	56.00
Nonstudents	104.42	132.81	56.77

Expected values (E)
(calculate assuming student type and # hrs are independent)

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Now we have O's and E's

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

Observed values (O)

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours
Community College Students	90.57	115.19	49.24
Four-Year College Students	103.00	131.00	56.00
Nonstudents	104.42	132.81	56.77

Expected values (E)
(calculate assuming student type and # hrs are independent)

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

$$\sum_{(i \cdot j)} \frac{(O - E)^2}{E}$$

- *O* = observed values
- *E* = expected values
- *i* = the number of rows in the table
- *j* = the number of columns in the table

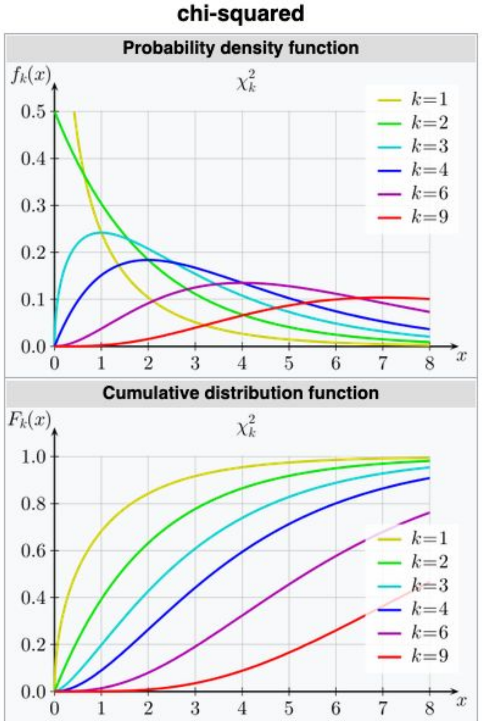
Now we calculate the test statistic:

□^{2} = $(111-90.57)^2$ / 90.57 + $(96-103)^2$/103 + ...

□^{2} = 12.99

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

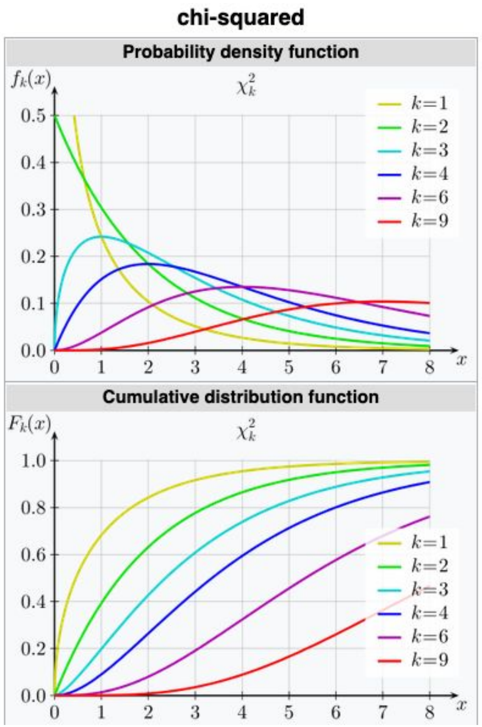
Is # hours volunteered independent of type of volunteer?



Need to compare this to a chi-squared distribution. The distribution itself has one parameter: *k*, the # of degrees of freedom

https://en.wikipedia.org/wiki/Chi-squared_distribution

Is # hours volunteered independent of type of volunteer?



Need to compare this to a chi-squared distribution. The distribution itself has one parameter: *k*, the # of degrees of freedom

degrees of freedom: (# rows - 1) * (# cols - 1)

https://en.wikipedia.org/wiki/Chi-squared_distribution

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

* When counting df, don't include total rows/cols

Need to compare this to a chi-squared distribution. The distribution itself has one parameter: *k*, the # of degrees of freedom

degrees of freedom: (# rows - 1) * (# cols - 1) = ?

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

Type of Volunteer	1–3 Hours	4–6 Hours	7–9 Hours	Row Total
Community College Students	111	96	48	255
Four-Year College Students	96	133	61	290
Nonstudents	91	150	53	294
Column Total	298	379	162	839

* When counting df, don't include total rows/cols

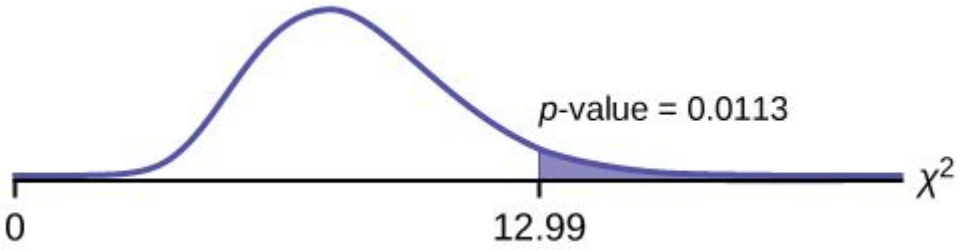
Need to compare this to a chi-squared distribution. The distribution itself has one parameter: *k*, the # of degrees of freedom

degrees of freedom: (# rows - 1) * (# cols - 1) = (3-1) * (3-1) = 2 * 2 = 4

Compare our test statistic against x^2_4

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

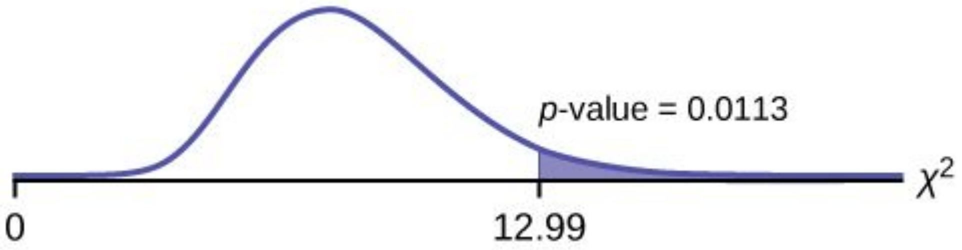


Is □^{2} = 12.99 spooky??

We can reject the null hypothesis (boring H_0= independent variables) for any alpha level > p-value of 0.0113

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

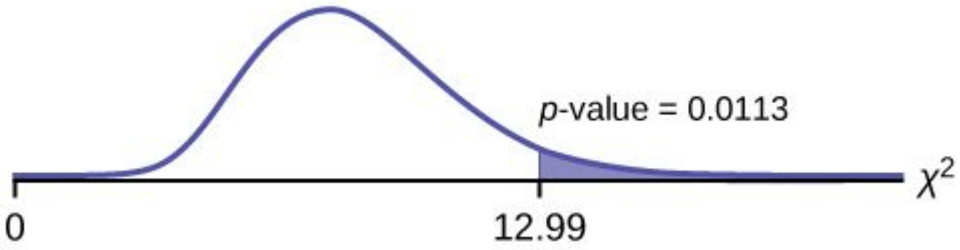


Is □^{2} = 12.99 spooky??

At the 5% level of significance, we reject the null hypothesis that volunteer type and # hours volunteered are independent.

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?

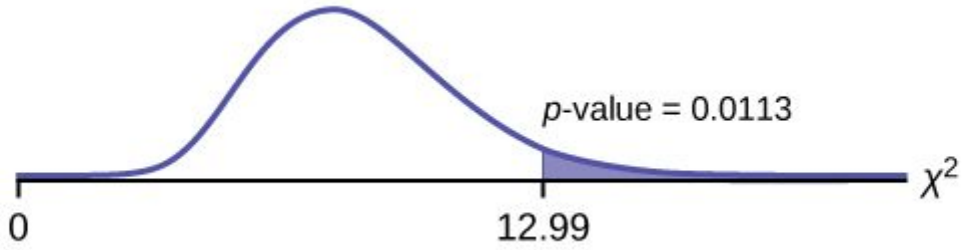


Is □^{2} = 12.99 spooky??

Does this necessitate that volunteer type and # hours volunteers are dependent on each other?

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Is # hours volunteered independent of type of volunteer?



Is □^{2} = 12.99 spooky??

Does this necessitate that volunteer type and # hours volunteers are dependent on each other? No, we just think it's pretty likely!

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Contingency Tables

- A = got a speeding ticket last year (binary)
- B = uses cell phone while driving (binary)

- 755 individuals are surveyed
    - 70 got a speeding ticket last year
    - 305 use cell phones while driving
- **What parts of a contingency table can you fill out?**

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Contingency Table

- 755 individuals are surveyed
  - 70 got a speeding ticket last year
  - 305 use cell phones while driving

- A = got a speeding ticket last year (binary)
- B = uses cell phone while driving (binary)

	A=0	A=1	Row Total
B=0			
B=1			
Column Total			

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Contingency Table

- 755 individuals are surveyed
  - 70 got a speeding ticket last year
  - 305 use cell phones while driving

- A = got a speeding ticket last year (binary)
- B = uses cell phone while driving (binary)

	A=0	A=1	Row Total
B=0	?	?	450 = 755-305
B=1	?	?	305
Column Total	685 = 755-70	70	755

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Calculating Expected Values

- A = got a speeding ticket last year (binary)
- B = uses cell phone while driving (binary)
- **What is E?**

	A=0	A=1	Row Total
B=0	?	?	O=450
B=1	?	E	O=305
Column Total	O=685	O=70	O=755

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Calculating Expected Values

- A = got a speeding ticket last year (binary)
- B = uses cell phone while driving (binary)
- E/755 = (70/755)*(305/755)
- E = 305*70/755 = row*col/total = 28.3

	A=0	A=1	Row Total
B=0	?	?	**O=450**
B=1	?	**E=28.3**	**O=305**
Column Total	**O=685**	**O=70**	**O=755**

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Interpreting Expected Values

- How do you interpret E= 28.3?
  - Reminder: E is the expected value under the null hypothesis
  - A = got a speeding ticket last year (binary)
  - B = uses cell phone while driving (binary)

	A=0	A=1	Row Total
B=0	?	?	O=450
B=1	?	E=28.3	O=305
Column Total	O=685	O=70	O=755

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Interpreting Expected Values

- **How do you interpret E= 28.3?**
    - If the null hypothesis is true (i.e., using a cell phone while driving and getting a speeding ticket are independent variables), we would expect ~28 people to use cell phones and get a speeding ticket

	A=0	A=1	Row Total
B=0	?	?	O=450
B=1	?	E=28.3	O=305
Column Total	O=685	O=70	O=755

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Recap: testing A and B independence

Steps to run a chi-squared test:

1. Have **observed O's** for orange cells
2. Calculate **expected E's** for orange cells

	A=0	A=1	Row Total
B=0	?	?	O=450
B=1	?	?	O=305
Column Total	O=685	O=70	O=755

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

Recap: testing A and B independence

Steps to run a chi-squared test:

1. Have **observed O's** for orange cells
2. Calculate **expected E's** for orange cells
3. Calculate **test statistic** (summing $(O-E)^2/E$ across orange cells)
4. Plot test statistic (vertical line) on **chi-squared distribution** (degrees of freedom = (2-1)*(2-1) = 1), decide if significant enough for you to make conclusions

	A=0	A=1	Row Total
B=0	?	?	O=450
B=1	?	?	O=305
Column Total	O=685	O=70	O=755

https://courses.lumenlearning.com/introstats1/chapter/test-of-independence/

1 min break & attendance

tinyurl.com/5des96n8

Admin

- Phase 2 feedback (not grades) will be released after class via Gradescope. If you don’t see any comments, please contact your mentor TA
- Phase 4/5 rubric will be discussed in Friday discussions this week (along with multiple hypothesis corrections)
- Phase 4 serves as the close-to-final draft, which you’ll get peer reviewed on; you can make edits to it for your final Phase 5 submission

Admin

- Your final project should be readable to a non-data-scientist (flow like an essay; interpret/explain everything; data cleaning code relegated to the appendix)
- Avoid “chart barf”: don’t just make plots of everything because you can. Make sure they serve a purpose (e.g. supporting / rejecting the data arguments you’re making re: your hypotheses)

Probabilities so far

Joint probability P(A, B) represents proportion of all combinations of values of variables A and B

Marginal probabilities P(A) and P(B) represent the summation of P(A, B) over all values of the "other" variable

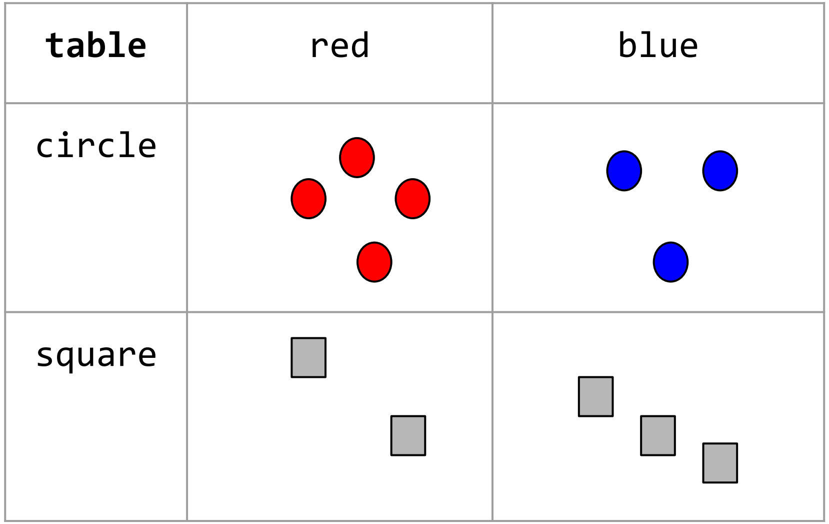
Probabilities so far

Joint probability P(A, B) represents proportion of all combinations of values of variables A and B

Marginal probabilities P(A) and P(B) represent the summation of P(A, B) over all values of the "other" variable

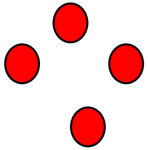
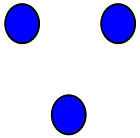
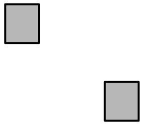
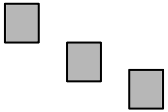
Conditional probabilities P(A | B) and P(B | A) each represent a "slice" of the joint probability table, renormalized

Conditional probability P(Color | shape) from a contingency table



given	red	blue
circle	?	?

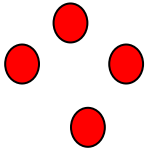
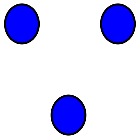
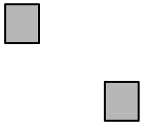
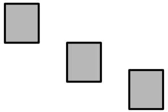
Conditional probability P(Color | shape) from a contingency table

table	red	blue
circle		
square		

given	red	blue
circle	4/7	?

Out of the 4+3 = 7 circles,
4 of them are red

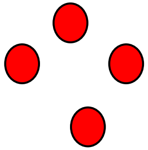
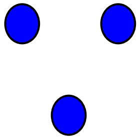
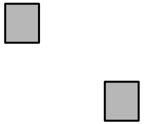
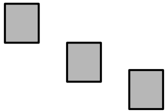
Conditional probability P(Color | shape) from a contingency table

table	red	blue
circle		
square		

given	red	blue
circle	4/7	?

What is P(blue | circle)?

Conditional probability P(Color | shape) from a contingency table

table	red	blue
circle		
square		

given	red	blue
circle	4/7	3/7

Out of the 4+3 = 7 circles,
3 of them are blue

Conditional probability from joint probability

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

marginal	red	blue
	1/2	1/2

Step 1: Calculate the marginal distribution of the conditioned (given) variable

Conditional probability from joint probability

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

Step 1: Calculate the marginal distribution of the conditioned (given) variable

Step 2: Divide each column of the joint probability by the value of the marginal for that column

Conditional probability from joint probability

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

=

given	red
circle	2/3
square	1/3

given	blue
circle	?
square	?

Step 1: Calculate the marginal distribution of the conditioned (given) variable

Step 2: Divide each column of the joint probability by the value of the marginal for that column

Pr(circle|red) = $\frac{1}{3}$ / $\frac{1}{2}$ = $\frac{2}{3}$
Pr(square|red) = $\frac{1}{6}$ / $\frac{1}{2}$ = $\frac{1}{3}$

Conditional probability from joint probability

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

=

given	red
circle	2/3
square	1/3

given	blue
circle	?
square	?

Step 1: Calculate the marginal distribution of the conditioned (given) variable

Step 2: Divide each column of the joint probability by the value of the marginal for that column

$\Pr(\text{circle}|\text{red}) = \frac{1}{3} / \frac{1}{2} = \frac{2}{3}$

$\Pr(\text{square}|\text{red}) = \frac{1}{6} / \frac{1}{2} = \frac{1}{3}$

Fill out these givens!

Conditional probability from joint probability

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

=

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

Step 1: Calculate the marginal distribution of the conditioned (given) variable

Step 2: Divide each column of the joint probability by the value of the marginal for that column

Pr(circle|red) = $\frac{1}{3}$ / $\frac{1}{2}$ = $\frac{2}{3}$
Pr(square|red) = $\frac{1}{6}$ / $\frac{1}{2}$ = $\frac{1}{3}$

Pr(circle|blue) = $\frac{1}{4}$ / $\frac{1}{2}$ = $\frac{1}{2}$
Pr(square|blue) = $\frac{1}{4}$ / $\frac{1}{2}$ = $\frac{1}{2}$

Divide joint by marginal to get conditionals

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

=

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

P(Shape | Color)

"Given"

Divide joint by marginal to get conditionals

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

=

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

$$\frac{P(Shape, Color)}{P(Color)}$$

=

P(Shape | Color)

"Given"

Divide joint by marginal to get conditionals

Is the joint probability

- greater than
- less than
- equal to

the conditional probability?

$$\frac{\texttt{P(Shape, Color)}}{\texttt{P(Color)}}$$

$$=$$

$$\texttt{P(Shape | Color)}$$

Divide joint by marginal to get conditionals

Is the joint probability

- ~~greater than~~
- **less than**
- equal to **only if P(Color) is 1**

the conditional probability?

$$\frac{\texttt{P(Shape, Color)}}{\texttt{P(Color)}}$$

$$=$$

$$\texttt{P(Shape | Color)}$$

Divide joint by marginal to get conditionals

Is the joint probability

- ~~greater than~~
- **less than**
- equal to **only if P(Color) is 1**

the conditional probability?

- **P(Shape,Color) = P(Color)*P(Shape|Color)**

$$\frac{P(\text{Shape, Color})}{P(\text{Color})} = P(\text{Shape} \mid \text{Color})$$

Divide joint by marginal to get conditionals

Is the joint probability

- ~~greater than~~
- **less than**
- equal to **only if P(Color) is 1**

the conditional probability?

$$\frac{\texttt{P(Shape, Color)}}{\texttt{P(Color)}} = \texttt{P(Shape | Color)}$$

- **P(Shape,Color) = P(Color)*P(Shape|Color)**
- **P(Color) is a probability P(Color) so <=1.**

Divide joint by marginal to get conditionals

Is the joint probability

$$\frac{P(\text{Shape, Color})}{P(\text{Color})}$$

- ~~greater than~~
- **less than**
- equal to **only if P(Color) is 1**

=

the conditional probability?

$$P(\text{Shape} \mid \text{Color})$$

- **P(Shape,Color) = P(Color)*P(Shape|Color)**
- **P(Color) is a probability P(Color) so <=1.**
- **Hence P(Shape,Color) <= P(Shape|Color)**

Divide joint by marginal to get conditionals

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	red	blue
	1/2	1/2

=

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

$$\frac{P(Shape, Color)}{P(Color)}$$

=

$$P(Shape \mid Color)$$

"Given"

Multiply conditionals by marginal to get joint

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

P(Shape, Color)

=

P(Shape | Color) P(Color)

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

Can you get back to the joint from the conditionals without the marginal?

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

Can you get back to the joint from the conditionals without the marginal?

No! There are an infinite number of different joint distributions that have the same conditionals

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

How many degrees of freedom does the joint distribution table have?

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

How many degrees of freedom does the joint distribution table have?

3 (remember, we can figure out the 4th value by doing 1 - sum(other three))!

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

DF for joint distribution: 3

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

DF for joint distribution: 3

DF for red conditionals: 1 (you can get the other value by subtracting from 1)

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

DF for joint distribution: 3

DF for red conditionals: 1 (you can get the other value by subtracting from 1)

DF for blue conditionals: 1 (you can get the other value by subtracting from 1)

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

DF for joint distribution: 3

DF for conditionals: 2 total

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

	red	blue
	-	-

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

DF for joint distribution: 3

This means we need 1 DF from the marginal in order to recover the 3 total DF needed for the joint distribution!

DF for conditionals: 2 total

Multiply conditionals by marginal to get joint

joint	red	blue
circle	?	?
square	?	?

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

This means we need 1 DF from the marginal in order to recover the 3 total DF needed for the joint distribution!

Hence, we cannot get back to the joint from the conditionals without the marginal.

Divide by marginal going the other way

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	?	?

P(red|square) P(blue|square)

Divide by marginal going the other way

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

P(red|square) = (⅙) / (5/12) = ⅖

P(blue|square) = (1/4) / (5/12) = 3/5

Divide by marginal going the other way

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

$$\frac{\boxed{P(\text{Shape}, \text{Color})}}{P(\text{Shape})} = P(\text{Color} \mid \text{Shape})$$

Divide by marginal going the other way

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

$$\frac{P(\text{Shape, Color})}{\boxed{P(\text{Shape})}} = P(\text{Color} \mid \text{Shape})$$

Divide by marginal going the other way

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

$$\frac{P(\text{Shape, Color})}{P(\text{Shape})} = \boxed{P(\text{Color} \mid \text{Shape})}$$

Divide by marginal going the other way

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

$$\frac{P(\text{Shape, Color})}{P(\text{Shape})} = P(\text{Color} \mid \text{Shape})$$

Can we express this in terms of a different conditional and marginal?

Bayes' Rule!

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

$$\frac{P(\text{Shape} \mid \text{Color})\, P(\text{Color})}{P(\text{Shape})}$$

$$= P(\text{Color} \mid \text{Shape})$$

Bayes' Rule!

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

$$\frac{\boxed{P(\text{Shape} \mid \text{Color})}\, P(\text{Color})}{P(\text{Shape})}$$

$$= P(\text{Color} \mid \text{Shape})$$

Bayes' Rule!

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

$$\frac{P(\text{Shape} \mid \text{Color})\ \boxed{P(\text{Color})}}{P(\text{Shape})}$$

$$= P(\text{Color} \mid \text{Shape})$$

Bayes' Rule!

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

given	blue
circle	1/2
square	1/2

$$\frac{P(\text{Shape} \mid \text{Color})\, P(\text{Color})}{\boxed{P(\text{Shape})}}$$

$$= P(\text{Color} \mid \text{Shape})$$

Bayes' Rule!

joint	red	blue
circle	1/3	1/4
square	1/6	1/4

/

marginal	
circle	7/12
square	5/12

=

given	red	blue
circle	4/7	3/7

given	red	blue
square	2/5	3/5

=

marginal	red	blue
	1/2	1/2

*

given	red
circle	2/3
square	1/3

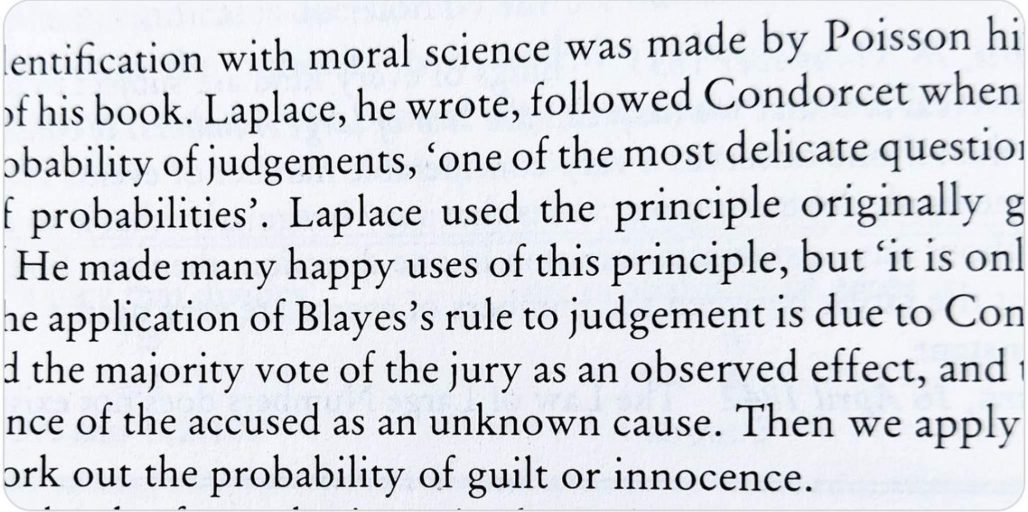
given	blue
circle	1/2
square	1/2

$$\frac{P(\text{Shape} \mid \text{Color})\, P(\text{Color})}{P(\text{Shape})}$$

$$= \boxed{P(\text{Color} \mid \text{Shape})}$$

Dean Eckles
@deaneckles

TIL that Poisson misspelled Rev. Thomas Bayes' name throughout his work in the 1830s



lentification with moral science was made by Poisson hi
of his book. Laplace, he wrote, followed Condorcet when
obability of judgements, 'one of the most delicate questio
f probabilities'. Laplace used the principle originally g
He made many happy uses of this principle, but 'it is onl
he application of Blayes's rule to judgement is due to Con
d the majority vote of the jury as an observed effect, and
nce of the accused as an unknown cause. Then we apply
ork out the probability of guilt or innocence.

11:09 AM · Oct 11, 2022 · Twitter for Android

https://twitter.com/deaneckles/status/1579851655298027520

What is Bayes' rule for?

We want to know about *causes* but we can usually only observe *effects*

It's much easier to estimate the probability of effects given causes

Bayes' rule lets us reverse the logic of causality!

Actual real-life urn model problem

Prof. Koenecke: "we had almost no trick-or-treaters, in the end one group of rowdy teenagers emptied out the bowl on our porch…"

Student on Nov 1st: "we found a few unattended bowls and emptied them out!"

Did this student raid the candy bowl?

Did the student raid the candy bowl?

Koenecke urn probability distribution:

Entirely M&Ms

https://www.walmart.com/ip/M-M-s-Milk-Chocolate-Candy-Full-Size-1-69-oz-Bag/38253480

Did the student raid the candy bowl?

Koenecke urn probability distribution:

Entirely M&Ms

Samples from unknown student urn distribution:

Lots of tootsie pops

https://www.walmart.com/ip/M-M-s-Milk-Chocolate-Candy-Full-Size-1-69-oz-Bag/38253480

https://www.amazon.com/Tootsie-Roll-Assorted-Pops-100/dp/B003TRYMHS

Did the student raid the candy bowl?

P(raiding Koenecke | mostly tootsie pops)

https://www.walmart.com/ip/M-M-s-Milk-Chocolate-Candy-Full-Size-1-69-oz-Bag/38253480

https://www.amazon.com/Tootsie-Roll-Assorted-Pops-100/dp/B003TRYMHS

Did the student raid the candy bowl?

P(raiding Koenecke | mostly tootsie pops)

What do we need to know?

The components of Bayes' rule!

https://www.walmart.com/ip/M-M-s-Milk-Chocolate-Candy-Full-Size-1-69-oz-Bag/38253480

https://www.amazon.com/Tootsie-Roll-Assorted-Pops-100/dp/B003TRYMHS

Did the student raid the candy bowl?

P(raiding Koenecke | mostly tootsie pops)

What do we need to know?

Conditional probability

P(tootsie pops | raiding Koenecke)

Prior probability

P(raiding Koenecke)

Marginal probability

P(tootsie pops)

https://www.walmart.com/ip/M-M-s-Milk-Chocolate-Candy-Full-Size-1-69-oz-Bag/38253480

https://www.amazon.com/Tootsie-Roll-Assorted-Pops-100/dp/B003TRYMHS

Did the student raid the candy bowl?

P(raiding Koenecke | mostly tootsie pops)

What do we need to know?

Conditional probability

P(tootsie pops | raiding Koenecke) is **low**

Prior probability

P(raiding Koenecke) is **moderate**

Marginal probability

P(tootsie pops) is **moderate**

https://www.walmart.com/ip/M-M-s-Milk-Chocolate-Candy-Full-Size-1-69-oz-Bag/38253480

https://www.amazon.com/Tootsie-Roll-Assorted-Pops-100/dp/B003TRYMHS

Did the student raid the candy bowl?

$$P(\text{raiding Koenecke} \mid \text{tootsie pops}) = \frac{P(\text{tootsie pops} \mid \text{raiding Koenecke})\ P(\text{raiding Koenecke})}{P(\text{tootsie pops})}$$

= some low number

What is Bayes' rule for?

We want to know about *causes* but we can usually only observe *effects*

Cause: which house? Effect: what candy?

It's much easier to estimate the probability of effects given causes

Bayes' rule lets us reverse the logic of causality!

What is Bayes' rule for?

- Lots of teaching examples involve drawing “balls from urns” because it’s an easy way to explain the concept
- But, there are lots of places you might use Bayes’ rule in data science jobs, e.g.: estimating probabilities about spam filters detecting spam emails, given they are spam (or not spam)
- This is why lots of data science interviews test for things like understanding Bayes’ rule!

Bayes' rule for spam

- ~50% of the emails you receive are spam emails
- Spam detection software claims that, given a spam email, it has a 90% probability of detecting it as spam
- Their probability for a false positive (given a non-spam email, it is detected as spam) is 10%

Bayes' rule for spam

- ~50% of the emails you receive are spam emails P(spam) = 0.5
- Spam detection software claims that, given a spam email, it has a 90% probability of detecting it as spam

- Their probability for a false positive (given a non-spam email, it is detected as spam) is 10%

Bayes' rule for spam

- ~50% of the emails you receive are spam emails
  P(spam) = 0.5
- Spam detection software claims that, given a spam email, it has a 90% probability of detecting it as spam
  P(detected as spam | spam) = 0.9
- Their probability for a false positive (given a non-spam email, it is detected as spam) is 10%

Bayes' rule for spam

- ~50% of the emails you receive are spam emails
  P(spam) = 0.5
- Spam detection software claims that, given a spam email, it has a 90% probability of detecting it as spam
  P(detected as spam | spam) = 0.9
- Their probability for a false positive (given a non-spam email, it is detected as spam) is 10%
  P(detected as spam | not spam) = 0.1

Bayes' rule for spam

- P(spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Given that an email is detected as spam, what is the probability that it's actually a non-spam email?

Bayes' rule for spam

- P(spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Given that an email is detected as spam, what is the probability that it's actually a non-spam email?

P(not spam | detected as spam)

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 1 - 0.5 = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50, P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes:

P(not spam | detected as spam)

= P(not spam, detected as spam) / P(detected as spam)

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50, P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes</u>:

P(not spam | detected as spam)

= P(not spam, detected as spam) / P(detected as spam)

= P(detected as spam | not spam) * P(not spam) / P(detected as spam)

= 0.1 * 0.5 / P(detected as spam)

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50, P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes:

P(not spam | detected as spam)

= P(not spam, detected as spam) / P(detected as spam)

= P(detected as spam | not spam) * P(not spam) / P(detected as spam)

= 0.1 * 0.5 / P(detected as spam)

How do we get this??

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50 → P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes Part 2:</u>

P(detected as spam)

= P(detected as spam | spam)*P(spam)

+ P(detected as spam | not spam)*P(not spam)

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes Part 2:</u>

P(detected as spam)

= P(detected as spam | spam)*P(spam)

+ P(detected as spam | not spam)*P(not spam)

Why?

- P(A|B)*P(B) = P(A, B)

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes Part 2:</u>

P(detected as spam)

= $\boxed{\text{P(detected as spam}}$ | spam)*P(spam)

+ P(detected as spam | not spam)*P(not spam)

Probability detected as spam and is truly spam

Why?

- P(A|B)*P(B) = P(A, B)

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes Part 2:</u>

P(detected as spam)

= P(detected as spam | spam)*P(spam)

+ $\boxed{\text{P(detected as spam | not spam)*P(not spam)}}$

Probability detected as spam but ISN'T truly spam

Why?

- P(A|B)*P(B) = P(A, B)
- P(A|1-B)*P(1-B) = P(A, 1-B)

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes Part 2:</u>

P(detected as spam)

= P(detected as spam | spam)*P(spam) + P(detected as spam | not spam)*P(not spam)

Probability email is detected as spam (whether or not it is truly spam)

Why?

- P(A|B)*P(B) = P(A, B)
- P(A|1-B)*P(1-B) = P(A, 1-B)
- P(A,B)+P(A,1-B) = P(A)

Bayes' rule for spam

New Goal: P(detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

<u>Bayes Part 2:</u>

P(detected as spam)

= P(detected as spam | spam)*P(spam)
  + P(detected as spam | not spam)*P(not spam)

= 0.9*0.5 + 0.1*0.5 = 0.5

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50, P(not spam) = 0.50
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes:

P(not spam | detected as spam)

= P(not spam, detected as spam) / P(detected as spam)

= P(detected as spam | not spam) * P(not spam) / P(detected as spam)

= 0.1 * 0.5 / P(detected as spam)

How do we get this??

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50, P(not spam) = 0.50
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Bayes:

P(not spam | detected as spam)

= P(not spam, detected as spam) / P(detected as spam)

= P(detected as spam | not spam) * P(not spam) / P(detected as spam)

= 0.1 * 0.5 / P(detected as spam) = 0.1 * 0.5 / 0.5 = 0.1

Bayes' rule for spam

Goal: P(not spam | detected as spam)

Known:

- P(spam) = 0.50→ P(not spam) = 0.5
- P(detected as spam | spam) = 0.9
- P(detected as spam | not spam) = 0.1

Using Bayes rule (twice) tells us: Given that an email is detected as spam, the probability that it's actually a non-spam email is 0.1.

https://mindfulmodeler.substack.com/p/bayesian-inference-from-first-principles