Conditional Probability

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
SLIDES ADAPTED FROM DAVE BLEI AND LAUREN HANNAH

Administrivia

- Autograder
- Office Hours

Context

- Data science is often worried about “if-then” questions
  - If my e-mail looks like this, is it spam?
  - If I buy this stock, will my portfolio improve?
- Since data science uses the language of probabilities, we need *conditional* probabilities (continuing probability intro)
- Also need to **combine** distributions

Conditional Probabilities

The *conditional probability* of event A given event B is the probability of A when B is known to occur,

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

Conditional Probabilities

The *conditional probability* of event A given event B is the probability of A when B is known to occur,

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

Conditional Probabilities

The *conditional probability* of event A given event B is the probability of A when B is known to occur,

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

Conditional Probabilities

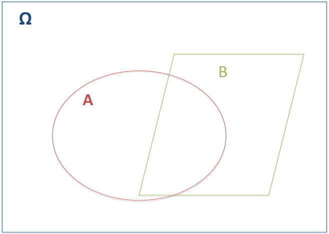
The *conditional probability* of event A given event B is the probability of A when B is known to occur,

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

Conditional Probabilities

The *conditional probability* of event A given event B is the probability of A when B is known to occur,

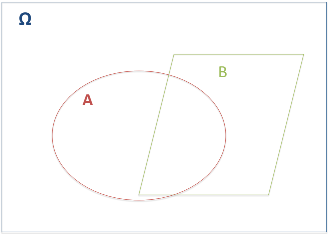
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

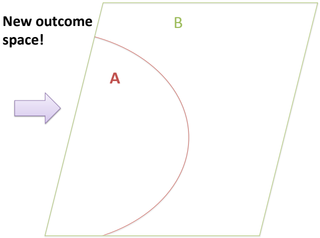


Conditional Probabilities

The *conditional probability* of event A given event B is the probability of A when B is known to occur,

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





Independence (Reminder)

Random variables X and Y are independent if and only if $P(X = x, Y = y) = P(X = x)P(Y = y)$. How does this interact with conditional probabilities?

Conditional probabilities equal unconditional probabilities with independence:

- $P(X = x \mid Y) = P(X = x)$
- *Knowing Y tells us nothing about X*

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

$$P(A > 3 \cap B + A = 6) =$$

$$P(A > 3) =$$

$$P(A > 3 | B + A = 6) =$$

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

$$P(A > 3 \cap B + A = 6) = \frac{2}{36}$$

$$P(A > 3) =$$

$$P(A > 3 \mid B + A = 6) =$$

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

$$P(A > 3 \cap B + A = 6) = \frac{2}{36}$$

$$P(A > 3) = \frac{3}{6}$$

$$P(A > 3 \mid B + A = 6) =$$

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

$$P(A > 3 \cap B + A = 6) = \frac{2}{36}$$

$$P(A > 3) = \frac{3}{6}$$

$$P(A > 3 \mid B + A = 6) = \frac{\frac{2}{36}}{\frac{3}{6}} = \frac{2}{36}\frac{6}{3}$$

Conditional Probabilities

Example

What is the probability that the sum of two dice is six given that the first is greater than three?

- $A \equiv$ First die
- $B \equiv$ Second die

	B=1	B=2	B=3	B=4	B=5	B=6
A=1	2	3	4	5	6	7
A=2	3	4	5	6	7	8
A=3	4	5	6	7	8	9
A=4	5	6	7	8	9	10
A=5	6	7	8	9	10	11
A=6	7	8	9	10	11	12

$$P(A > 3 \cap B + A = 6) = \frac{2}{36}$$

$$P(A > 3) = \frac{3}{6}$$

$$P(A > 3 \mid B + A = 6) = \frac{\frac{2}{36}}{\frac{3}{6}} = \frac{2}{36}\frac{6}{3}$$

$$= \frac{1}{9}$$

Combining Distributions

- Somtimes distributions you have aren't what you need
  - Conditional $\longrightarrow$ joint (chain)
  - Reverse conditional direction (Bayes')

The chain rule

- The definition of conditional probability lets us derive the *chain rule*, which let's us define the joint distribution as a product of conditionals:

$$P(X, Y) = P(X, Y)\frac{P(Y)}{P(Y)}$$

The chain rule

- The definition of conditional probability lets us derive the *chain rule*, which let's us define the joint distribution as a product of conditionals:

$$
\begin{aligned}
P(X, Y) &= P(X, Y)\frac{P(Y)}{P(Y)} \\
&= P(X|Y)P(Y)
\end{aligned}
$$

The chain rule

- The definition of conditional probability lets us derive the *chain rule*, which let's us define the joint distribution as a product of conditionals:

$$
\begin{aligned}
P(X,Y) &= P(X,Y)\frac{P(Y)}{P(Y)} \\
&= P(X|Y)P(Y)
\end{aligned}
$$

- For example, let Y be a disease and X be a symptom. We may know $P(X|Y)$ and $P(Y)$ from data. Use the chain rule to obtain the probability of having the disease and the symptom.
- In general, for any set of N variables

$$
P(X_1,\ldots,X_N) = \prod_{n=1}^{N} P(X_n|X_1,\ldots,X_{n-1})
$$

Bayes' Rule

What is the relationship between $P(A|B)$ and $P(B|A)$?

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

1. Start with $P(A|B)$
2. Change outcome space from B to Ω
3. Change outcome space again from Ω to A

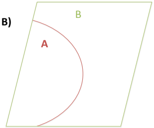
Bayes' Rule

What is the relationship between $P(A|B)$ and $P(B|A)$?

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

1. Start with $P(A|B)$
2. Change outcome space from B to Ω
3. Change outcome space again from Ω to A

P(A|B)

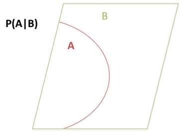


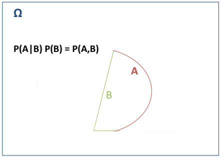
Bayes' Rule

What is the relationship between $P(A|B)$ and $P(B|A)$?

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

1. Start with $P(A|B)$
2. Change outcome space from B to Ω: $P(A|B)P(B)$
3. Change outcome space again from Ω to A





Bayes' Rule

What is the relationship between $P(A|B)$ and $P(B|A)$?

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

1. Start with $P(A|B)$
2. Change outcome space from B to Ω: $P(A|B)P(B)$
3. Change outcome space again from Ω to A: $\frac{P(A|B)P(B)}{P(A)}$