Probabilities and Data

Digging into Data: Jordan Boyd-Graber

University of Maryland

February 4, 2013

COLLEGE OF INFORMATION STUDIES

Slides adapted from Dave Blei and Lauren Hannah

Roadmap

- What are probabilities
  - Discrete
  - Continuous
- How to manipulate probabilities
- Properties of probabilities

Preface: Why make us do this?

- Probabilities are the language we use to describe data
- A reasonable (but geeky) definition of data science is how to get probabilities we care about from data
- Later classes will be about how to do this for different probability models and different types of data
- But first, we need key definitions of probability

Preface: Why make us do this?

- Probabilities are the language we use to describe data
- A reasonable (but geeky) definition of data science is how to get probabilities we care about from data
- Later classes will be about how to do this for different probability models and different types of data
- But first, we need key definitions of probability
- So pay attention!

Preface: Why make us do this?

- Probabilities are the language we use to describe data
- A reasonable (but geeky) definition of data science is how to get probabilities we care about from data
- Later classes will be about how to do this for different probability models and different types of data
- But first, we need key definitions of probability
- So pay attention!
- Also, ya'll need to get your environments set up

Outline

Card problem (from David MacKay)

- There are three cards
  - Red/Red
  - Red/Black
  - Black/Black

Card problem (from David MacKay)

- There are three cards
  - Red/Red
  - Red/Black
  - Black/Black
- I go through the following process.
  - Close my eyes and pick a card
  - Pick a side at random
  - Show you that side

Card problem (from David MacKay)

- There are three cards
  - Red/Red
  - Red/Black
  - Black/Black
- I go through the following process.
  - Close my eyes and pick a card
  - Pick a side at random
  - Show you that side
- Suppose I show you red. What's the probability the other side is red too? (Write down your answer!)

Random variable

- Probability is about *random variables.*
- A random variable is any "probabilistic" outcome.
- For example,
  - The flip of a coin
  - The height of someone chosen randomly from a population
- We'll see that it's sometimes useful to think of quantities that are not strictly probabilistic as random variables.
  - The temperature on 11/12/2013
  - The temperature on 03/04/1905
  - The number of times "streetlight" appears in a document

Random variable

- Random variables take on values in a *sample space*.
- They can be *discrete* or *continuous*:
  - Coin flip: $\{H, T\}$
  - Height: positive real values $(0, \infty)$
  - Temperature: real values $(-\infty, \infty)$
  - Number of words in a document: Positive integers $\{1, 2, \ldots\}$
- We call the outcomes *events*.
- Denote the random variable with a capital letter; denote a realization of the random variable with a lower case letter.
- E.g., X is a coin flip, x is the value (H or T) of that coin flip.

Discrete distribution

- A discrete distribution assigns a probability to every event in the sample space
- For example, if X is an (unfair) coin, then

$$\begin{aligned} P(X=H) &= 0.7 \\ P(X=T) &= 0.3 \end{aligned}$$

- The probabilities over the entire space must sum to one

$$\sum_x P(X=x) = 1$$

- And probabilities have to be greater than 0
- Probabilities of disjunctions are sums over part of the space. E.g., the probability that a die is bigger than 3:

$$P(D>3) = P(D=4) + P(D=5) + P(D=6)$$

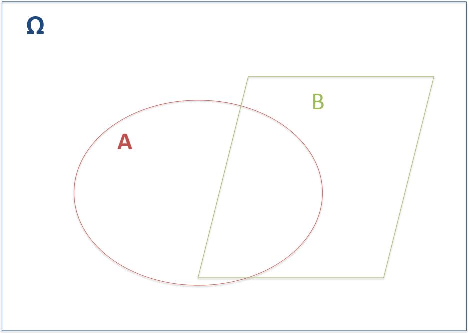
Outline

Events

An *event* is a set of outcomes to which a probability is assigned, for example, getting a card with Red on both sides.

Intersections and unions:

- Intersection: $P(A \cap B)$
- Union: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



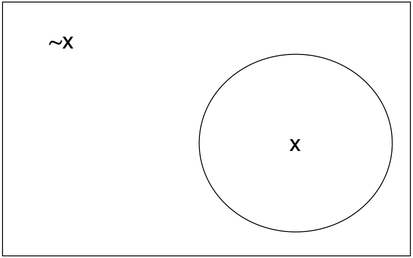
Joint distribution

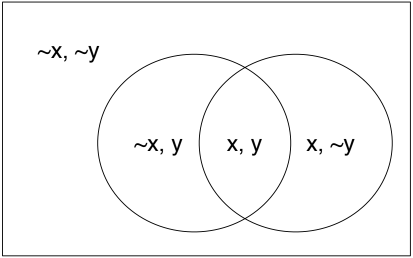
- Typically, we consider collections of random variables.
- The joint distribution is a distribution over the configuration of all the random variables in the ensemble.
- For example, imagine flipping 4 coins. The joint distribution is over the space of all possible outcomes of the four coins.

$$
\begin{aligned}
P(HHHH) &= 0.0625 \\
P(HHHT) &= 0.0625 \\
P(HHTH) &= 0.0625 \\
&\dots
\end{aligned}
$$

- You can think of it as a single random variable with 16 values.

Visualizing a joint distribution





Marginalization

If we are given a joint distribution, what if we are only interested in the distribution of one of the variables?

We can compute the distribution of $P(X)$ from $P(X, Y, Z)$ through *marginalization*:

$$\begin{aligned}\sum_y \sum_z P(X, Y = y, Z = z) &= \sum_y \sum_z P(X) P(Y = y, Z = z | X) \\ &= P(X) \sum_y \sum_z P(Y = y, Z = z | X) \\ &= P(X)\end{aligned}$$

Marginalization (from Leyton-Brown)

Joint distribution

temperature (T) and weather (W)

	T=Hot	T=Mild	T=Cold
W=Sunny	.10	.20	.10
W=Cloudy	.05	.35	.20

Marginalization allows us to compute distributions over smaller sets of variables:

- $P(X,Y) = \sum_z P(X,Y,Z=z)$
- Corresponds to summing out a table dimension
- New table still sums to 1

- Marginalize out weather
- Marginalize out temperature

Marginalization (from Leyton-Brown)

Joint distribution

temperature (T) and weather (W)

	T=Hot	T=Mild	T=Cold
W=Sunny	.10	.20	.10
W=Cloudy	.05	.35	.20

Marginalization allows us to compute distributions over smaller sets of variables:

- $P(X,Y) = \sum_z P(X,Y,Z=z)$
- Corresponds to summing out a table dimension
- New table still sums to 1

- Marginalize out weather

T=Hot	T=Mild	T=Cold
.15	.55	.30

- Marginalize out temperature

Marginalization (from Leyton-Brown)

Joint distribution

temperature (T) and weather (W)

	T=Hot	T=Mild	T=Cold
W=Sunny	.10	.20	.10
W=Cloudy	.05	.35	.20

Marginalization allows us to compute distributions over smaller sets of variables:

- $P(X,Y) = \sum_z P(X,Y,Z=z)$
- Corresponds to summing out a table dimension
- New table still sums to 1

- Marginalize out weather

T=Hot	T=Mild	T=Cold
.15	.55	.30

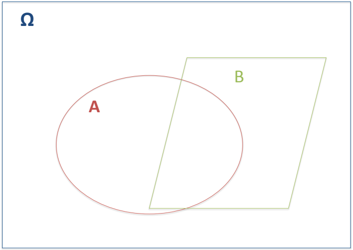
- Marginalize out temperature

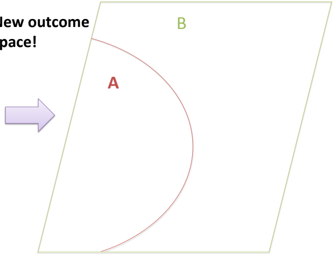
W=Sunny	.40
W=Cloudy	.60

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

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

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

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

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

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

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

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

Outline

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

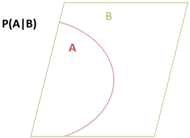
$$P(X_1, \ldots, X_N) = \prod_{n=1}^{N} P(X_n | X_1, \ldots, X_{n-1})$$

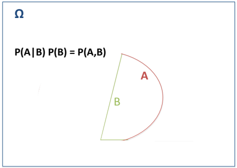
Bayes' Rule

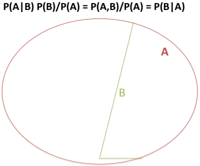
What is the relationship between $P(A|B)$ and $P(B|A)$?

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

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







Independence

Random variables X and Y are independent if and only if $P(X = x, Y = y) = P(X = x)P(Y = y)$.

Conditional probabilities equal unconditional probabilities with independence:

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

Mathematical examples:

- If I draw two socks from my (multicolored) laundry, is the color of the first sock independent from the color of the second sock?
- If I flip a coin twice, is the first outcome independent from the second outcome?

Independence

Intuitive Examples:

- Independent:
  - you use a Mac / the Green line is on schedule
  - snowfall in the Himalayas / your favorite color is blue
- Not independent:
  - you vote for Mitt Romney / you are a Republican
  - there is a traffic jam on the Beltway / the Redskins are playing

Independence

Are these independent?

- the values of two dice
- the value of the first die and the sum of the values
- whether it is raining and the number of taxi cabs
- whether it is raining and the amount of time it takes me to hail a cab
- the first two words in a sentence

Independence

Example: two coins, C_1, C_2 with

$$P(H \mid C_1) = 0.5, \qquad P(H \mid C_2) = 0.3$$

Suppose that I randomly choose a number $Y \in \{1, 2\}$ and take coin C_Z. I flip it twice, with results (X_1, X_2)

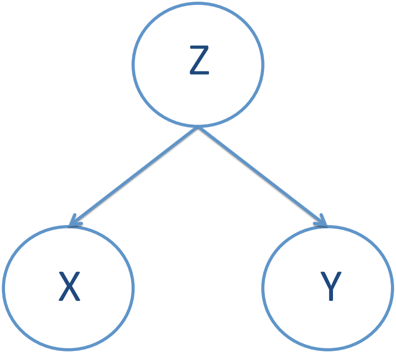
- are X_1 and X_2 independent?
- what about if I know Y?

Conditional Independence

Two random variables (or events) X and Y are *conditionally independent* given Z if and only if

$$P(X = x, Y = y|Z) = P(X = x|Z)P(Y = y|Z)$$

Graphical model notation:



Outline

Returning to the card problem

- Now we can solve the card problem.
- Let X_1 be the random side of the random card I chose
- Let X_2 be the other side of that card
- Compute $P(X_2 = \mathrm{red} | X_1 = \mathrm{red})$

$$P(X_2 = \mathrm{red} | X_1 = \mathrm{red}) = \frac{P(X_1 = R, X_2 = R)}{P(X_1 = R)} \tag{1}$$

Returning to the card problem

- Now we can solve the card problem.
- Let X_1 be the random side of the random card I chose
- Let X_2 be the other side of that card
- Compute $P(X_2 = \text{red} | X_1 = \text{red})$

$$P(X_2 = \text{red} | X_1 = \text{red}) = \frac{P(X_1 = R, X_2 = R)}{P(X_1 = R)} \tag{1}$$

- Numerator is 1/3: Only one card has two red sides.
- Denominator is 1/2: There are three possible sides of the six that are red.

Conditional Probabilities

I have 5 socks in my dryer: 3 gray, 2 blue.

- pull one out
- pull second one out
- possible outcomes: GG, GB, BG, BB
- Socks:
  $P(2nd\ Sock\ G \mid 1st\ Sock\ G) =$
- Dice: I roll 2 dice, look at total:
  $P(Total = 7 \mid Total \leq 7) =$

Conditional Probabilities

I have 5 socks in my dryer: 3 gray, 2 blue.

- pull one out
- pull second one out
- possible outcomes: GG, GB, BG, BB
- Socks:
  $P(2nd\ Sock\ G \mid 1st\ Sock\ G) = \frac{3/5 \cdot 2/4}{3/5} = \frac{1}{2}$
- Dice: I roll 2 dice, look at total:
  $P(Total\ = 7 \mid Total\ \leq 7) =$

Conditional Probabilities

I have 5 socks in my dryer: 3 gray, 2 blue.

- pull one out
- pull second one out
- possible outcomes: GG, GB, BG, BB

- Socks:
  $P(2nd\ Sock\ G \mid 1st\ Sock\ G) = \frac{3/5 \cdot 2/4}{3/5} = \frac{1}{2}$
- Dice: I roll 2 dice, look at total:
  $P(Total\ = 7 \mid Total\ \leq 7) = \frac{6/36}{21/36} = \frac{6}{21}$

Joint Probability Distributions

A *joint probability distribution* is a probability distribution over a set of random variables.

Example: let X be the color of the first sock, Y the color of the second. I have 5 socks in my dryer: 3 gray, 2 blue.

		Y G	Y B
X	G		
X	B		

Can extend conditional probabilities to conditional distributions: $P(Y \mid X = G)$ or $P(X \mid Y = B)$

Joint Probability Distributions

A *joint probability distribution* is a probability distribution over a set of random variables.

Example: let X be the color of the first sock, Y the color of the second. I have 5 socks in my dryer: 3 gray, 2 blue.

		Y	
		G	B
X	G	$3/5 \cdot 2/4 = 3/10$	
	B		

Can extend conditional probabilities to conditional distributions: $P(Y \mid X = G)$ or $P(X \mid Y = B)$

Joint Probability Distributions

A *joint probability distribution* is a probability distribution over a set of random variables.

Example: let X be the color of the first sock, Y the color of the second. I have 5 socks in my dryer: 3 gray, 2 blue.

		Y	
		G	B
X	G	$3/5 \cdot 2/4 = 3/10$	$3/5 \cdot 2/4 = 3/10$
	B		

Can extend conditional probabilities to conditional distributions: $P(Y \mid X = G)$ or $P(X \mid Y = B)$

Joint Probability Distributions

A *joint probability distribution* is a probability distribution over a set of random variables.

Example: let X be the color of the first sock, Y the color of the second. I have 5 socks in my dryer: 3 gray, 2 blue.

		Y	
		G	B
X	G	$3/5 \cdot 2/4 = 3/10$	$3/5 \cdot 2/4 = 3/10$
	B	$2/5 \cdot 3/4 = 3/10$	

Can extend conditional probabilities to conditional distributions: $P(Y \mid X = G)$ or $P(X \mid Y = B)$

Joint Probability Distributions

A *joint probability distribution* is a probability distribution over a set of random variables.

Example: let X be the color of the first sock, Y the color of the second. I have 5 socks in my dryer: 3 gray, 2 blue.

		Y	
		G	B
X	G	$3/5 \cdot 2/4 = 3/10$	$3/5 \cdot 2/4 = 3/10$
	B	$2/5 \cdot 3/4 = 3/10$	$2/5 \cdot 1/4 = 1/10$

Can extend conditional probabilities to conditional distributions: $P(Y \mid X = G)$ or $P(X \mid Y = B)$

Outline

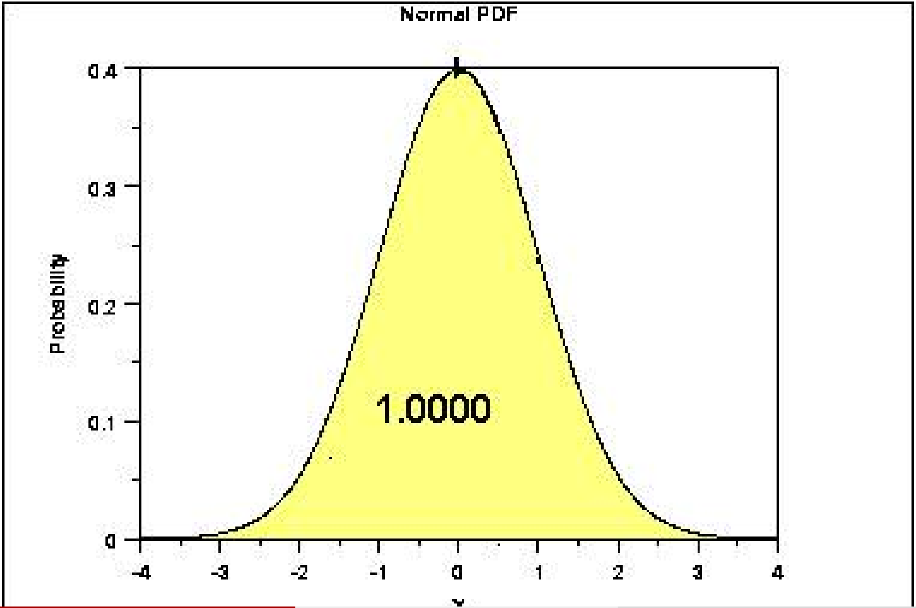
Continuous random variables

- We've only used discrete random variables so far (e.g., dice)
- Random variables can be continuous.
- We need a *density* $p(x)$, which *integrates* to one.
  E.g., if $x \in \mathbb{R}$ then
  $$\int_{-\infty}^{\infty} p(x)dx = 1$$
- Probabilities are integrals over smaller intervals. E.g.,
  $$P(X \in (-2.4, 6.5)) = \int_{-2.4}^{6.5} p(x)dx$$
- Notice when we use P, p, X, and x.

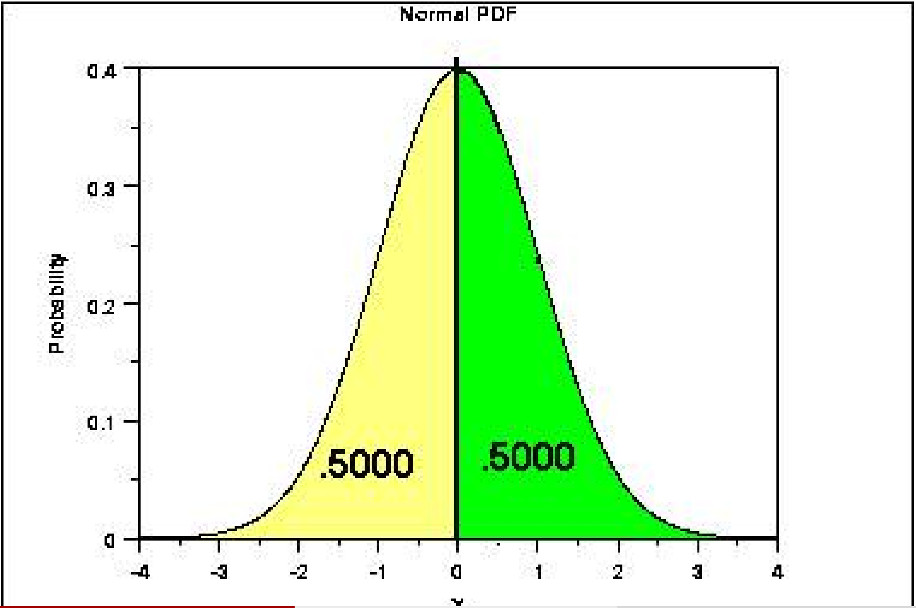
Continuous random variables

- We've only used discrete random variables so far (e.g., dice)
- Random variables can be continuous.
- We need a *density* $p(x)$, which *integrates* to one.
  E.g., if $x \in \mathbb{R}$ then
  $$\int_{-\infty}^{\infty} p(x)dx = 1$$
- Probabilities are integrals over smaller intervals. E.g.,
  $$P(X \in (-2.4, 6.5)) = \int_{-2.4}^{6.5} p(x)dx$$
- Notice when we use P, p, X, and x.
- Integrals? I didn't sign up for this!

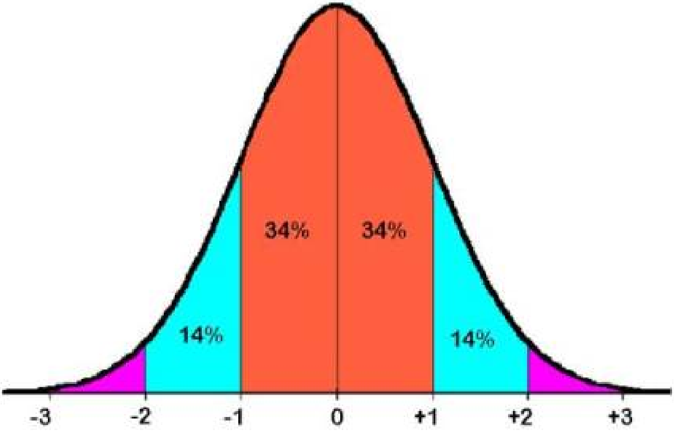
Integrals?



Integrals?



Integrals?



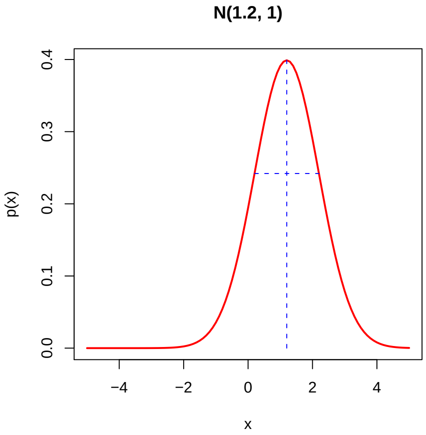
The Gaussian distribution

- The Gaussian (or Normal) is a continuous distribution.

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- The density of a point x is proportional to the negative exponentiated half distance to μ scaled by σ^2.
- μ is called the *mean*; σ^2 is called the *variance*.

Gaussian density



- The mean μ controls the location of the bump.
- The variance σ^2 controls the spread of the bump.

Outline

Expectation

An *expectation* of a random variable is a weighted average:

$$
\begin{aligned}
\mathrm{E}[f(X)] &= \sum_{x=1}^{\infty} f(x)\, p(x) && \text{(discrete)} \\
&= \int_{-\infty}^{\infty} f(x)\, p(x)\, dx && \text{(continuous)}
\end{aligned}
$$

Alternate formulation for positive random variables:

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{x=1}^{\infty} P(X > x) && \text{(discrete)} \\
&= \int_{0}^{\infty} P(X > x)\, dx && \text{(continuous)}
\end{aligned}
$$

Expectation

Expectations of constants or known values:

- $\mathrm{E}[a] = a$
- $\mathrm{E}[Y \mid Y = y] = y$

Expectation

Example: Gaussian distribution $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&=
\end{aligned}
$$

Expectation

Example: Gaussian distribution $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} \mathrm{E}[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \mu \end{aligned}$$

Expectation of die / dice

What is the expectation of the roll of die?

Expectation of die / dice

What is the expectation of the roll of die?

One die

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} =$$

Expectation of die / dice

What is the expectation of the roll of die?

One die

$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$

Expectation of die / dice

What is the expectation of the roll of die?

One die

$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$

What is the expectation of the sum of two dice?

Expectation of die / dice

What is the expectation of the roll of die?

One die

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

What is the expectation of the sum of two dice?

Two die

$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} =$$

Expectation of die / dice

What is the expectation of the roll of die?

One die

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

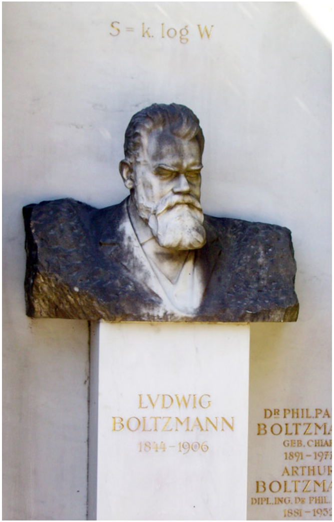
What is the expectation of the sum of two dice?

Two die

$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7$$

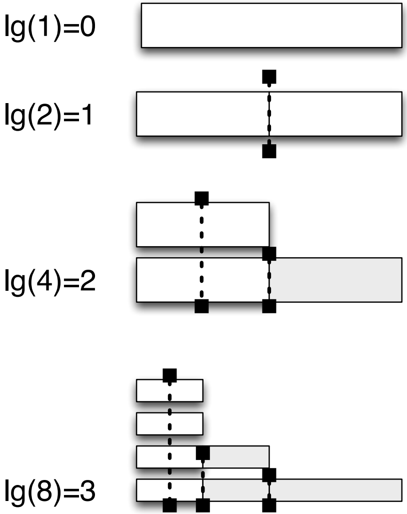
Entropy

- Measure of disorder in a system
- In the real world, entroy in a system tends to increase
- Can also be applied to probabilities:
  - Is one (or a few) outcomes certain (low entropy)
  - Are things equiprobable (high entropy)
- In data science
  - We look for features that allow us to *reduce* entropy (decision trees)
  - All else being equal, we seek models that have *maximum* entropy (Occam's razor)



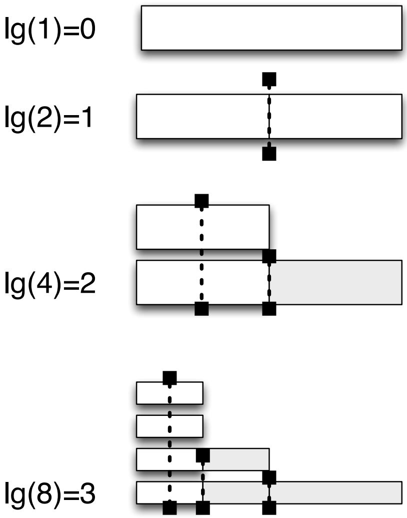
Aside: Logarithms

- $\lg(x) = b \Leftrightarrow 2^b = x$
- Makes big numbers small
- Way to think about them: cutting a carrot



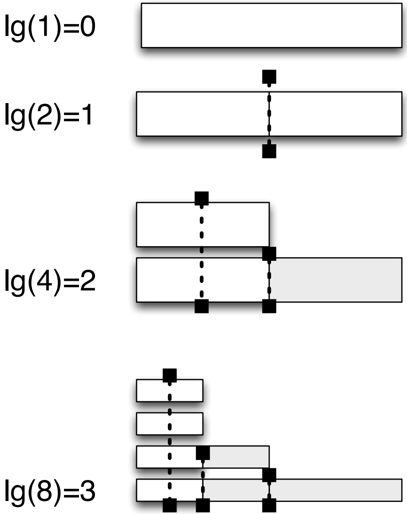
Aside: Logarithms

- $\lg(x) = b \Leftrightarrow 2^b = x$
- Makes big numbers small
- Way to think about them: cutting a carrot
- Negative numbers?



Aside: Logarithms

- $\lg(x) = b \Leftrightarrow 2^b = x$
- Makes big numbers small
- Way to think about them: cutting a carrot
- Negative numbers?
- Non-integers?



Entropy

Entropy is a measure of uncertainty that is associated with the distribution of a random variable:

$$\begin{aligned} H(X) &= -\mathrm{E}\left[\lg(p(X))\right] \\ &= -\sum_x p(x)\lg(p(x)) && \text{(discrete)} \\ &= -\int_{-\infty}^{\infty} p(x)\lg(p(x))\,dx && \text{(continuous)} \end{aligned}$$

Does not account for the values of the random variable, only the spread of the distribution.

- $H(X) \geq 0$
- uniform distribution = highest entropy, point mass = lowest
- suppose $P(X=1)=p,\ P(X=0)=1-p$ and $P(Y=100)=p,\ P(Y=0)=1-p$: X and Y have the same entropy

Entropy of a die / dice

What is the entropy of a roll of a die?

Entropy of a die / dice

What is the entropy of a roll of a die?

One die

$$-\left(\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)\right)=2.58$$

Entropy of a die / dice

What is the entropy of a roll of a die?

One die

$$-\left(\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)+\frac{1}{6}\lg\left(\frac{1}{6}\right)\right)=2.58$$

What is the entropy of the sum of two die? Tricky question: will it be higher or lower than the first one?

Entropy of a die / dice

What is the entropy of a roll of a die?

One die

$$-\left(\tfrac{1}{6}\lg\left(\tfrac{1}{6}\right)+\tfrac{1}{6}\lg\left(\tfrac{1}{6}\right)+\tfrac{1}{6}\lg\left(\tfrac{1}{6}\right)+\tfrac{1}{6}\lg\left(\tfrac{1}{6}\right)+\tfrac{1}{6}\lg\left(\tfrac{1}{6}\right)+\tfrac{1}{6}\lg\left(\tfrac{1}{6}\right)\right)=2.58$$

What is the entropy of the sum of two die? Tricky question: will it be higher or lower than the first one?

Two die

$$\begin{aligned}-\Bigg(\frac{1}{36}\lg\left(\frac{1}{36}\right)&+\frac{2}{36}\lg\left(\frac{2}{36}\right)+\frac{3}{36}\lg\left(\frac{3}{36}\right)+\frac{4}{36}\lg\left(\frac{4}{36}\right)+\frac{5}{36}\lg\left(\frac{5}{36}\right)\\&+\frac{6}{36}\lg\left(\frac{6}{36}\right)+\frac{5}{36}\lg\left(\frac{5}{36}\right)+\frac{4}{36}\lg\left(\frac{4}{36}\right)+\frac{3}{36}\lg\left(\frac{3}{36}\right)\\&+\frac{2}{36}\lg\left(\frac{2}{36}\right)+\frac{1}{36}\lg\left(\frac{1}{36}\right)\Bigg)=3.27\end{aligned}$$

Whew!

- That's it for now
- You don't have to be an expert on this stuff (there are other classes for that)
- This is to get your feet wet and to know the concepts when you see the math

First assignment

- Find some data
- Find interesting relationships in your data (next week!)
- Use Rattle to display those relationships (be creative and thorough!)