Expectations and Entropy

Data Science: Jordan Boyd-Graber
University of Maryland
SLIDES ADAPTED FROM DAVE BLEI AND LAUREN HANNAH

Expectation

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

$$\mathrm{E}[f(X)] = \sum_x f(x)\,p(x) \qquad \text{(discrete)}$$

$$= \int_{-\infty}^{\infty} f(x)\,p(x)\,dx \qquad \text{(continuous)}$$

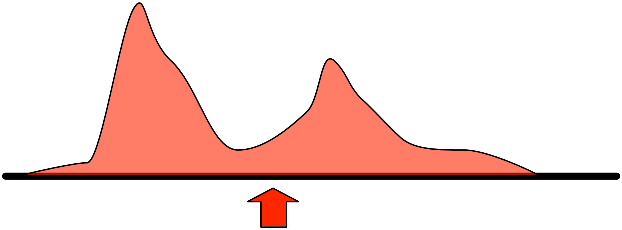
Expectation

Expectations of constants or known values:

- $\mathrm{E}[a] = a$

Expectation Intuition

- $\mathbb{E}[x]$ is most common expectation
- Average outcome (might not be an event: 2.4 children)
- Center of mass



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

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

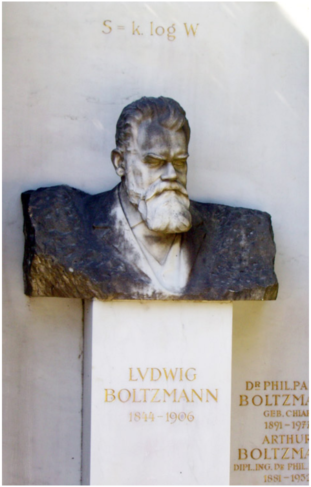
What is the expectation of the sum of two dice?

Two die

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

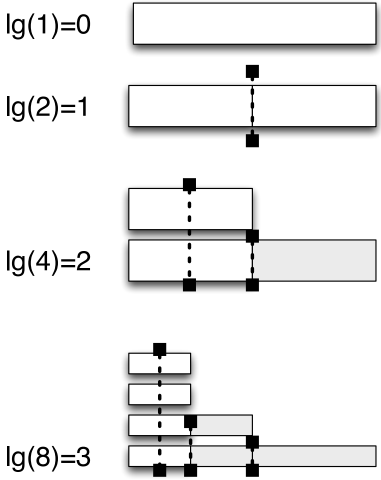
Entropy

- 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 <u>reduce</u> entropy (decision trees)
  - All else being equal, we seek models that have <u>maximum</u> entropy (Occam's razor)



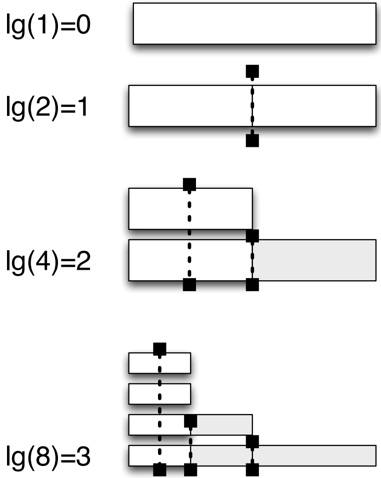
Aside: Logarithms

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



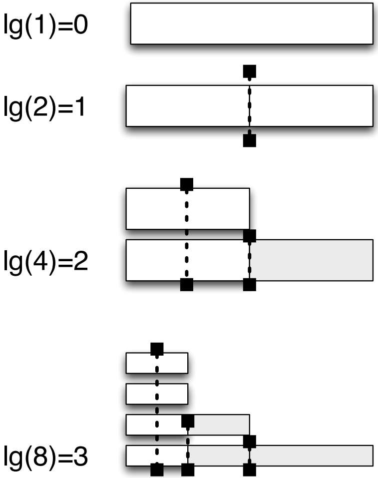
Aside: Logarithms

- $\lg(x) = b \Leftrightarrow 2^b = x$
- 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:

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

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

Wrap up

- Probabilities are the language of data science
- You'll need to manipulate probabilities and understand marginalization and independence
- In Class: Working through probability examples
- Next: **Conditional** probabilities