Willem van den Boom

March 29, 2016

Discuss in groups of 2 or 3:

How to quantify the information obtained by knowing the value x of a random variable X?

Equivalently:
How to quantify the uncertainty induced by not knowing the value of X?

Today's Topic: ENTROPY

Willem van den Boom

March 29, 2016

Learning goals

- Definition of entropy
- Explore its properties based on examples
- An example of an application of entropy to statistical inference

The definition of entropy H

- Quantify the information obtained by knowing the value x of a random variable X
  Equivalently, quantify the uncertainty induced by not knowing the value of X

The definition of entropy H

- Quantify the information obtained by knowing the value x of a random variable X
  Equivalently, quantify the uncertainty induced by not knowing the value of X
- For a discrete random variable X with probability mass function $p(x)$ and support S, its entropy is given by

$$H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} = \mathbb{E}\left[\log \frac{1}{p(X)}\right].$$

Example: Degenerate distribution

What is the entropy if $p(0) = 1$, $\mathbb{P}[X = 0] = 1$, and $S = \{0\}$?

$$H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} =$$

Example: Degenerate distribution

What is the entropy if $p(0) = 1$, $\mathbb{P}[X = 0] = 1$, and $S = \{0\}$?

$$H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} = \log \frac{1}{1} = 0$$

If X is deterministic, knowing its value provides no information.

Can entropy be negative?

Example: Uniform distribution

What is the entropy of $p(x) = \frac{1}{|S|}$, the discrete uniform distribution?

$$H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} =$$

Example: Uniform distribution

What is the entropy of $p(x) = \frac{1}{|S|}$, the discrete uniform distribution?

$$H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} = \log \frac{1}{p(x)} = \log |S|$$

With logarithmic base 2, this is the minimal length of a binary code that captures all information in the event $X = x$.

Example: Binary entropy

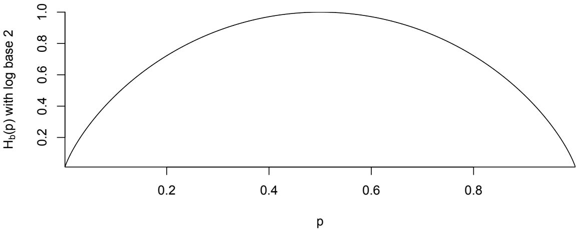
What is the entropy if $S = \{0, 1\}$, $p(1) = p$ and $p(0) = 1 - p$?

$$H_b(p) = H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} =$$

Example: Binary entropy

What is the entropy if $S = \{0, 1\}$, $p(1) = p$ and $p(0) = 1 - p$?

$$H_b(p) = H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$



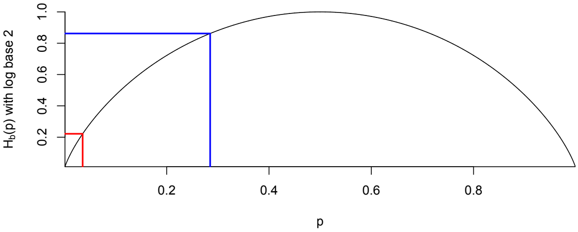
What has higher entropy: The random variable that it rains on September 1 in Durham or in Death Valley?

Durham has on average 104 days with precipitation per year (365 days) and Death Valley 13 days.

Example: Binary entropy

What is the entropy if $S = \{0, 1\}$, $p(1) = p$ and $p(0) = 1 - p$?

$$H_b(p) = H(X) = \sum_{x \in S} p(x) \log \frac{1}{p(x)} = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$



What has higher entropy: The random variable that it rains on September 1 in Durham or in Death Valley?

Durham has on average 104 days with precipitation per year and Death Valley 13 days.

Entropy in statistical inference

- Conditional entropy: Consider X and Y with joint pmf $p(x, y) = p(x \mid y)p(y)$, then the conditional entropy is

$$H(X \mid Y) = \sum_{x,y} p(x, y) \log \frac{1}{p(x \mid y)} = \mathbb{E}\left[\log \frac{1}{p(X \mid Y)}\right].$$

Entropy in statistical inference

- Conditional entropy: Consider X and Y with joint pmf $p(x, y) = p(x \mid y)p(y)$, then the conditional entropy is

$$H(X \mid Y) = \sum_{x,y} p(x, y) \log \frac{1}{p(x \mid y)} = \mathbb{E}\left[\log \frac{1}{p(X \mid Y)}\right].$$

- Fano's inequality: For any estimator $\hat{X}(Y)$ of X where X has support S,

$$\mathbb{P}\left[\hat{X}(Y) \neq X\right] \geq \frac{H(X \mid Y) - 1}{\log |S|}.$$

Entropy in statistical inference

- Conditional entropy: Consider X and Y with joint pmf $p(x, y) = p(x \mid y)p(y)$, then the conditional entropy is

$$H(X \mid Y) = \sum_{x,y} p(x, y) \log \frac{1}{p(x \mid y)} = \mathbb{E}\left[\log \frac{1}{p(X \mid Y)}\right].$$

- Fano's inequality: For any estimator $\hat{X}(Y)$ of X where X has support S,

$$\mathbb{P}\left[\hat{X}(Y) \neq X\right] \geq \frac{H(X \mid Y) - 1}{\log |S|}.$$

When is the bound on the right close to 1?

Learning goals

- Definition of entropy
- Explore its properties based on examples
- An example of an application of entropy to statistical inference

Questions?