Quantifying Randomness
Part 1: Introduction to Entropy

INFO-1301, Quantitative Reasoning 1

University of Colorado Boulder

October 3, 2016

Prof. Michael Paul

Prof. William Aspray

How uncertain is a distribution?

One extreme: everything is equally likely

$P(X=1) = 0.2$
$P(X=2) = 0.2$
$P(X=3) = 0.2$
$P(X=4) = 0.2$
$P(X=5) = 0.2$

With this distribution, you are completely uncertain about what the outcome will be

How uncertain is a distribution?

Another extreme: only one outcome is likely

$P(X=1) = 0.0$
$P(X=2) = 0.0$
$P(X=3) = 1.0$
$P(X=4) = 0.0$
$P(X=5) = 0.0$

With this distribution, you are completely certain about what the outcome will be

How uncertain is a distribution?

Distributions in between these extremes:

$P(X=1) = 0.1$
$P(X=2) = 0.2$
$P(X=3) = 0.4$
$P(X=4) = 0.2$
$P(X=5) = 0.1$

$P(X=1) = 0.05$
$P(X=2) = 0.05$
$P(X=3) = 0.8$
$P(X=4) = 0.05$
$P(X=5) = 0.05$

Which one is more “certain”?

- Which one would be easier to base decisions on?

Information Entropy

Entropy is a measurement of how evenly distributed a probability distribution is

- We can quantify the “uncertainty” of a distribution, using the concepts from the preceding examples

Entropy of a random variable *X* is denoted H(*X*)

- Entropy is non-negative (0 or higher)

Lower entropy means it is less even, more certain
Higher entropy means it is more even, less certain

Information Entropy

The lowest possible value of entropy is 0

This occurs when a distribution gives 0 probability to all but one outcome (called a **point** distribution)

$P(X=1) = 0.0$	$P(X=1) = 1.0$
$P(X=2) = 0.0$	$P(X=2) = 0.0$
$P(X=3) = 1.0$	$P(X=3) = 0.0$
$P(X=4) = 0.0$	$P(X=4) = 0.0$
$P(X=5) = 0.0$	$P(X=5) = 0.0$

Information Entropy

The highest possible value of entropy occurs when the distribution is **uniform**

$P(X=1) = 0.2$
$P(X=2) = 0.2$
$P(X=3) = 0.2$
$P(X=4) = 0.2$
$P(X=5) = 0.2$

Logarithms

If: $2^y = x$ then: $\log_2(x) = y$

↑

2 is the **base** of the log
Entropy uses base 2

$\log_2(2) = 1$

$\log_2(4) = 2$

$\log_2(8) = 3$

How many times can you divide by 2 until you get 1?

16 / 2 = 8 / 2 = 4 / 2 = 2 / 2 = 1

- ➢ $\log_2(16) = 4$

Logarithms

If: $2^y = x$ then: $\log_2(x) = y$

$\log_2(1) = 0$	$2^0 = 1$
$\log_2(10) = 3.32$	Between $\log_2(8)$ and $\log_2(16)$
$\log_2(1.5) = 0.58$	Between $\log_2(1)$ and $\log_2(2)$
$\log_2(0.5) = -1$	$2^{-1} = ½ = 0.5$
$\log_2(0.2) = -2.32$	If $x < 1$, then $\log(x) < 0$
$\log_2(0) =$ n/a	If $x = 0$, there is no $\log(x)$
$\log_2(-0.5) =$ n/a	If $x < 0$, there is no $\log(x)$

Information Entropy

How to calculate H(*X*)?

For every outcome of *X*, calculate:

$P(X{=}a) \times \log_2 P(X{=}a)$

Then sum the results for each outcome:

$P(X{=}a) \times \log_2 P(X{=}a) + P(X{=}b) \times \log_2 P(X{=}b)$

Then multiply the final result by –1:

$- P(X{=}a) \times \log_2 P(X{=}a) - P(X{=}b) \times \log_2 P(X{=}b)$

Information Entropy

Example: coin flip

$P(X = \text{Heads}) = 0.5$
$P(X = \text{Tails}) = 0.5$

$0.5 \log_2(0.5) + 0.5 \log_2(0.5) = -1$

$H(X) = 1$

The highest possible entropy for a distribution with two outcomes is 1

Information Entropy

Example: chance of rain

P(*X* = Sunny) = 0.9
P(*X* = Rainy) = 0.1

$0.9 \log_2(0.9) + 0.1 \log_2(0.1) = -0.469$

H(X) = 0.469

Entropy of this distribution is lower than a coin flip

Information Entropy

General formula:

$H(X) = -\sum_a P(X = a) \log_2 P(X = a)$

Next time we'll look at where this formula comes from