

Introduction to Information Entropy

Based on: An Intuitive Explanation of the Information Entropy of a Random Variable by Daniel Shawcross Wilkerson <http://danielwilkerson.com/entropy.html>

Learning Objectives

After completion of this module the student will be able to

1. define and use random variables
2. explain the meaning of expected value, calculate the expected value of a random variable
3. understand the definition of information entropy

Knowledge and Skills

1. random variables
2. expected value of a random variable
3. information entropy

Prerequisites

1. calculating the probability of events
2. logarithm, base 2, base 10
3. graphing quadratic functions

Introduction

In-class Activity 1

1. *Play a round of Twenty Questions*

What are good questions in this game? How about asking questions like: "Is it the red table?" How can we do better?

2. *Now play a few rounds of the following game: One person thinks of a whole number between 0 and 100. The others should try to guess this number, asking yes/no questions. What are good questions in this game? Why? How much information is in a question? Which question has more information:*

Q1: *Is it 53?*

Q2: *Is it less than 11?*

Why?

In this unit we are going to answer these questions by looking at random variables and expected values of random variables. We will define information and the information entropy, and briefly discuss the history, significance and applications of this concept.

Random Variables

Consider our second game, and let the mystery number be denoted by A . We know that A can be any whole number 1-100. Think about questions Q1 and Q2, and let us introduce the following definitions

$$X = \begin{cases} 1 & \text{if } A = 53 \\ 0 & \text{if } A \neq 53, \end{cases}$$

and

$$Y = \begin{cases} 1 & \text{if } A < 11 \\ 0 & \text{if } A \geq 11. \end{cases}$$

X and Y represent abstract models for an “experiment” or “observation”. We want to find out something (the mystery number), and we ask a question/design an experiment. What can happen? The answer to our question (the result of our experiment) can be ‘yes’, which we denote with 1, or ‘no’, which we denote with 0. We call X and Y *random variables*. Thus $X = 1$ corresponds to ‘the number is 53’, while $X = 0$ corresponds to ‘the number is not 53’. We say that the possible values of X and Y are 0 and 1.

The formal definition of random variable involves the notion of “probability space” - which you could think of as all the possible states of the ‘world’, in this case, all the possible subsets of the numbers 1-100. Then random variables can be simply defined as functions on the probability space.

In-class Activity 2

Roll a regular die once. Define at least three different random variables for this experiment.

Random variables and their expected value

In-class Activity 3

Suppose a lottery ticket costs \$1.00. The rules state that one in a thousand tickets wins \$100.00. How much do you expect to “win” if you play this lottery for a long time?

When you find the long-time average of a random variable, you approximate its expected value. Formally, the expected value of a random variable X with possible values $\{x_1, \dots, x_N\}$ is

$$E(X) = \sum_{j=1}^N x_j \cdot p(x_j).$$

So to compute it, follow the following steps:

1. consider each possible value of the random variable $X : x_i$
2. compute the probabilities $p(X = x_i)$
3. calculate the products $x_i \cdot p(X = x_i)$
4. add-up these terms

So for the above activity let our random variable X represent the two possibilities: winning \$99.00 or losing \$1.00, which we could say amounts to winning \$-1.00. What are the corresponding probabilities? Since one in a thousand tickets wins, winning \$99.00 has probability $\frac{1}{1000}$. This implies that losing a dollar has probability $\frac{999}{1000}$. Thus the expected value of X is

$$E(X) = 99 \cdot \frac{1}{1000} + (-1) \cdot \frac{999}{1000} = \frac{99}{1000} - \frac{999}{1000} = -\frac{900}{1000} = -0.9.$$

What does this mean? Is it worth buying these lottery tickets?

Homework Problems

1. How many winning tickets should there be for this lottery to be profitable to the players?
2. Consider a lottery in your state. How much do you expect to “win” on this lottery? (data collection)

Random variables and information

How much information is preserved in a random variable? Suppose the only possible value of X is 5. To observe this random variable, and see that it is 5, is not very surprising. It doesn't contain any information, so its information content should be 0. Now what if X takes the value 5 90% of the time, and 3 10% of the time. To see a 5 now, should have somewhat more information value, and the way we define it is

$$\log_2 \frac{1}{p(X=5)} = \log_2 \frac{1}{9/10} = -\log_2 \frac{9}{10} = 0.152.$$

In general, the information value of x_i is $\log_2 p(x_i)$.

Information entropy

Information entropy H of a random variable is defined as the expected information in X . More precisely, if X is a random variable with possible values $\{x_1, \dots, x_N\}$, then

$$H(X) = \sum_{j=1}^N p(x_j) \cdot \log_2 p(x_j).$$

Let's go back to Twenty Questions. Remember, X there is a random variable that is 1 if $A = 53$, and 0 if A is not 53. What is the information entropy of this random variable? How much do we gain by asking this question? We'll need the probabilities here. If A is a randomly chosen number, then the probability of $A = 53$ (and of $X = 1$) is $\frac{1}{100}$. This leaves $p(X = 0) = \frac{99}{100}$. To compute $H(X)$ we take

$$H(X) = -(0.01 \cdot \log_2 0.01 + 0.99 \cdot \log_2 0.99) = 0.0808.$$

(Note that the actual values of the random variable are not important, we only need their respective probabilities.) Now let's calculate the entropy of question Q2. What is the probability that A is less than 11? Exactly $\frac{10}{100}$. Thus,

$$H(Y) = -(0.1 \cdot \log_2 0.1 + 0.9 \cdot \log_2 0.9) = 0.469.$$

So what would be the best question to ask? Now we have a good way to measure the information content of a question.

In-class Activity 4

Think of tossing a coin, which is not necessarily fair, i.e., the probability of it coming up heads is p , where $0 \leq p \leq 1$. Identify a random variable, and graph its information entropy for the different p -s. How does this confirm your intuition about the best question to ask?

History, Connections and Applications

Information theory and information entropy was introduced by Claude Shannon in his 1948 paper “The Mathematical Theory for Communication”. Why entropy? The anecdote says that Shannon was reluctant to use the word “entropy” for this quantity, but Von Neumann told him to go ahead ”since nobody knows what entropy is, in a debate you will be sure to have an advantage. Shannon’s measure of entropy closely resembles the formulas used in thermodynamics. In statistical thermodynamics entropy S of a thermodynamics system is defined as

$$S = -k_B \sum p_i \log p_i,$$

where k_B is the Boltzmann constant ($\approx 1.38 \times 10^{-23} JK^{-1}$), and p_i is the probability microstate E_i occurs during the system’s fluctuations.

Links: information entropy: constant thermodynamics entropy: changing according to the second law of thermodynamics

Jaynes: thermodynamics could be seen as an application of Shannon’s information theory: the thermodynamic entropy is interpreted as being an estimate of the amount of further Shannon information needed to define the detailed microscopic state of the system, that remains uncommunicated by a description solely in terms of the macroscopic variables of classical thermodynamics. For example, adding heat to a system increases its thermodynamic entropy because it increases the number of possible microscopic states that it could be in, thus making any complete state description longer.

Extra activity Play the game ENTROPY

This game pits CHAOS, the force of time and decay against ORDER, the external constructive energy source, to destruct or build patterns.

Instructions (from Mind Sports Worldwide)

One player is Order, the other Chaos. Order is trying to make patterns

vertically and horizontally. Chaos is trying to prevent this. A pattern is any sequence of counters that is identical if viewed from either direction (left to right/right to left or up to down/down to up). A pattern scores the number of counters in the pattern. All patterns within a pattern also score. For example: red-green-blue-green-red scores 5 and also 3 for green-blue-green; total: 8. Another example: red-green-red-green-red scores 5 plus $3 \times 3 = 14$. A further example: red-red-red-red scores $4 + 2 \times 3 + 3 \times 2 = 16$. All possible patterns are listed below.

Chaos puts the counters in the bag. He draws them out unseen one at a time and places them on any empty square. Each time Chaos places a counter, Order may slide any one counter on the board (including the one just placed) vertically or horizontally over any number of vacant squares (exactly like a rook in chess). Only one counter may occupy a square. When the board is full every vertical and horizontal line is scored. The players then reverse roles and the player with the highest score is the winner.