

2

Probability, Entropy, and Inference

This chapter, and its sibling, Chapter 8, devote some time to notation. Just as the White Knight distinguished between the song, the name of the song, and what the name of the song was called (Carroll, 1998), we will sometimes need to be careful to distinguish between a random variable, the value of the random variable, and the proposition that asserts that the random variable has a particular value. In any particular chapter, however, I will use the most simple and friendly notation possible, at the risk of upsetting pure-minded readers. For example, if something is ‘true with probability 1’, I will usually simply say that it is ‘true’.

► 2.1 Probabilities and ensembles

An ensemble X is a triple $(x, \mathcal{A}_X, \mathcal{P}_X)$, where the *outcome* x is the value of a random variable, which takes on one of a set of possible values, $\mathcal{A}_X = \{a_1, a_2, \dots, a_i, \dots, a_I\}$, having probabilities $\mathcal{P}_X = \{p_1, p_2, \dots, p_I\}$, with $P(x = a_i) = p_i$, $p_i \geq 0$ and $\sum_{a_i \in \mathcal{A}_X} P(x = a_i) = 1$.

The name \mathcal{A} is mnemonic for ‘alphabet’. One example of an ensemble is a letter that is randomly selected from an English document. This ensemble is shown in figure 2.1. There are twenty-seven possible letters: a–z, and a space character ‘-’.

Abbreviations. Briefer notation will sometimes be used. For example, $P(x = a_i)$ may be written as $P(a_i)$ or $P(x)$.

Probability of a subset. If T is a subset of \mathcal{A}_X then:

$$P(T) = P(x \in T) = \sum_{a_i \in T} P(x = a_i). \quad (2.1)$$

For example, if we define V to be vowels from figure 2.1, $V = \{a, e, i, o, u\}$, then

$$P(V) = 0.06 + 0.09 + 0.06 + 0.07 + 0.03 = 0.31. \quad (2.2)$$

A joint ensemble XY is an ensemble in which each outcome is an ordered pair x, y with $x \in \mathcal{A}_X = \{a_1, \dots, a_I\}$ and $y \in \mathcal{A}_Y = \{b_1, \dots, b_J\}$.

We call $P(x, y)$ the joint probability of x and y .

Commas are optional when writing ordered pairs, so $xy \Leftrightarrow x, y$.

N.B. In a joint ensemble XY the two variables are not necessarily independent.

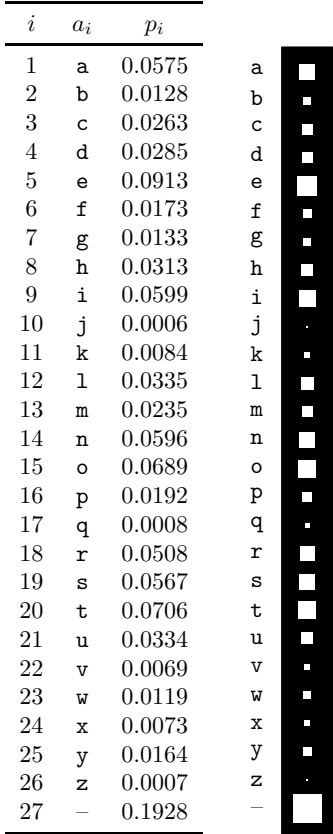


Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from *The Frequently Asked Questions Manual for Linux*). The picture shows the probabilities by the areas of white squares.