5 Probability

$$I = \sum_i p(A_i) \log_2 \left(\frac{1}{p(A_i)} \right)$$

为什么是log2？：

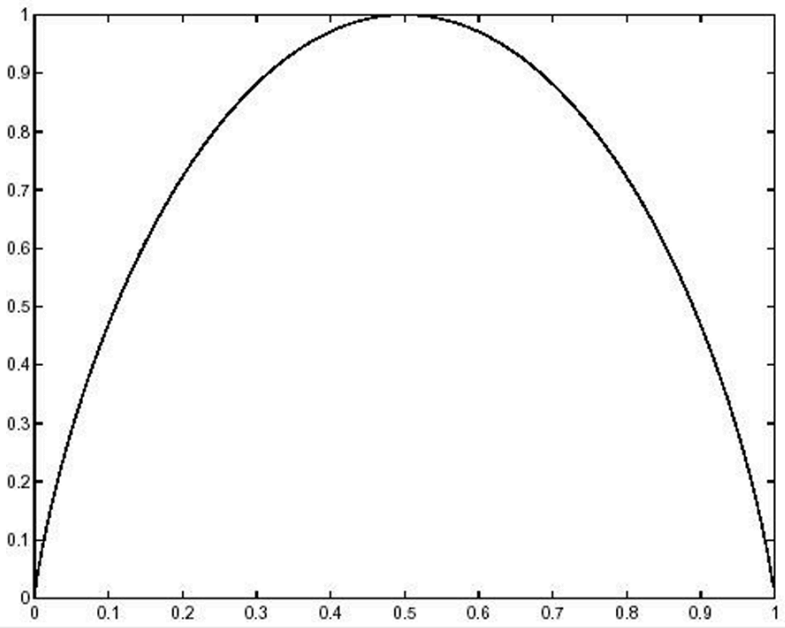
With base-2 logarithms the information is expressed in bits. Later, we will find natural logarithms to be useful.

If there are two events in the partition with probabilities p and $(1-p)$, the information per symbol is

$$I = p \log_2 \left(\frac{1}{p} \right) + (1-p) \log_2 \left(\frac{1}{1-p} \right) \tag{5.16}$$

which is shown, as a function of p, in Figure 5.3. It is largest (1 bit) for $p = 0.5$. Thus the information is a maximum when the probabilities of the two possible events are equal. Furthermore, for the entire range of probabilities between $p = 0.4$ and $p = 0.6$ the information is close to 1 bit. It is equal to 0 for $p = 0$ and for $p = 1$. This is reasonable because for such values of p the outcome is certain, so no information is gained by learning it.

For partitions with more than two possible events the information per symbol can be higher. If there are n possible events the information per symbol lies between 0 and $\log_2(n)$ bits, the maximum value being achieved when all probabilities are equal.



完了，本章节别的基本都学过。