Table 1.1: The joint distribution p(x, y) for two binary variables x and y.

1 Questions

- 1. Consider two binary variables x and y having the joint distribution given in Table 1.1. Compute the marginal distributions (p(x)) and p(y) and conditional distributions (p(y|x)) and p(x|y) from the table.
- 2. Evaluate the following quantities

(a)
$$H[x]$$
 (c) $H[y|x]$ (e) $H[x, y]$ (b) $H[y]$ (d) $H[x|y]$ (f) $I[x, y]$.

Use the formulae for entropy to compute the values required: leave them in the form of arithmetic combinations of logarithms rather than writing down the floating point results.

$$H(x) = -\sum_{i} p(x_i) \ln p(x_i)$$
(1.1)

$$H(x|y) = -\sum_{i} \sum_{j} p(x_i, y_j) \ln p(x_i|y_j)$$
 (1.2)

and similar definitions for H(y) and H(y|x)

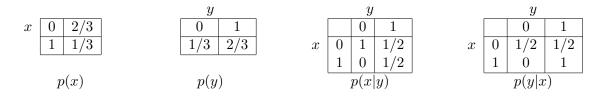
3. A fair coin is flipped until the first head occurs. Let X denote the number of flips required. Find the entropy H(X) in bits (i.e. using \log_2). This expression may be useful:

$$\sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2} \tag{1.3}$$

4. What is the minimum value of $H(p_1, ..., p_n) = H(\mathbf{p})$ as \mathbf{p} ranges over the set of n-dimensional probability vectors? Find all vectors \mathbf{p} which achieve this minimum. Discuss how this relates to the maximum entropy discrete distribution.

$\mathbf{2}$ Answers

1. From Table 1.1 we obtain the marginal probabilities by summation and the conditional probabilities by normalization, to give



2. From these tables for the first question, together with the definitions

$$H(x) = -\sum_{i} p(x_i) \ln p(x_i)$$
(2.1)

$$H(x) = -\sum_{i} p(x_{i}) \ln p(x_{i})$$

$$H(x|y) = -\sum_{i} \sum_{j} p(x_{i}, y_{j}) \ln p(x_{i}|y_{j})$$
(2.1)

and similar definitions for H(y) and H(y|x), we obtain the following results

(a)
$$H(x) = \ln 3 - \frac{2}{3} \ln 2$$

(b)
$$H(y) = \ln 3 - \frac{2}{3} \ln 2$$

(c)
$$H(y|x) = \frac{2}{3} \ln 2$$

(d)
$$H(x|y) = \frac{2}{3} \ln 2$$

(e)
$$H(x,y) = \ln 3$$

(f)
$$I(x;y) = \ln 3 - \frac{4}{3} \ln 2$$

3. The probability distribution for the random variable X is given by $P(X=i)=0.5^i$. Hence,

$$H(X) = -\sum_{i} p_{i} \log p_{i}$$

$$= -\sum_{i} 0.5^{i} \log(0.5^{i})$$

$$= -\log(0.5) \sum_{i} i \cdot 0.5^{i}$$

$$= \frac{0.5}{(1 - 0.5)^{2}}$$

$$= 2$$

4. Since $H(\mathbf{p}) \geq 0$ and $\sum_i p_i = 1$, then the minimum value for $H(\mathbf{p})$ is 0 which is achieved when $p_i = 1$ and $p_j = 0$ for $j \neq i$. The maximum entropy distribution is one that is uniform, so $p_i = p_j$ for all i, j. This is the opposite in that it is maximally peaked.