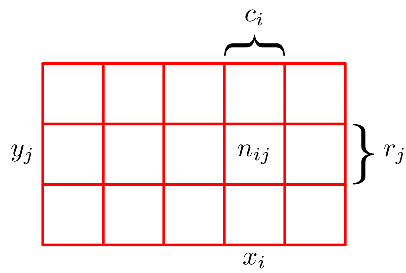

Figure 2.4 We can derive the sum and product rules of probability by considering a random variable X, which takes the values $\{x_i\}$ where $i = 1, \ldots, L$, and a second random variable Y, which takes the values $\{y_j\}$ where $j = 1, \ldots, M$. In this illustration, we have $L = 5$ and $M = 3$. If we consider the total number N of instances of these variables, then we denote the number of instances where $X = x_i$ and $Y = y_j$ by n_{ij}, which is the number of instances in the corresponding cell of the array. The number of instances in column i, corresponding to $X = x_i$, is denoted by c_i, and the number of instances in row j, corresponding to $Y = y_j$, is denoted by r_j.

given by the fraction of the total number of points that fall in column i, so that

$$p(X = x_i) = \frac{c_i}{N}. \tag{2.2}$$

Since $\sum_i c_i = N$, we see that

$$\sum_{i=1}^{L} p(X = x_i) = 1 \tag{2.3}$$

and, hence, the probabilities sum to one as required. Because the number of instances in column i in Figure 2.4 is just the sum of the number of instances in each cell of that column, we have $c_i = \sum_j n_{ij}$ and therefore, from (2.1) and (2.2), we have

$$p(X = x_i) = \sum_{j=1}^{M} p(X = x_i, Y = y_j), \tag{2.4}$$

which is the *sum rule* of probability. Note that $p(X = x_i)$ is sometimes called the *marginal* probability and is obtained by marginalizing, or summing out, the other variables (in this case Y).

If we consider only those instances for which $X = x_i$, then the fraction of such instances for which $Y = y_j$ is written $p(Y = y_j|X = x_i)$ and is called the *conditional* probability of $Y = y_j$ given $X = x_i$. It is obtained by finding the fraction of those points in column i that fall in cell i,j and, hence, is given by

$$p(Y = y_j|X = x_i) = \frac{n_{ij}}{c_i}. \tag{2.5}$$

Summing both sides over j and using $\sum_j n_{ij} = c_i$, we obtain

$$\sum_{j=1}^{M} p(Y = y_j|X = x_i) = 1 \tag{2.6}$$