Probability

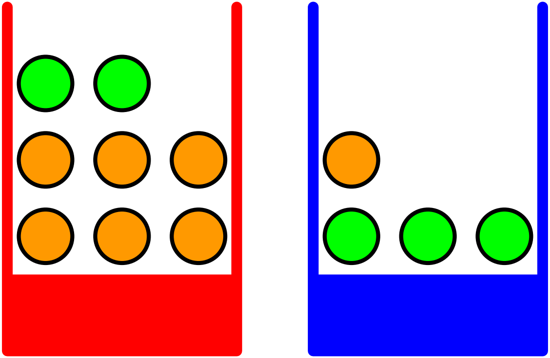
Martha Lewis

Department of Engineering Mathematics
University of Bristol

Probability

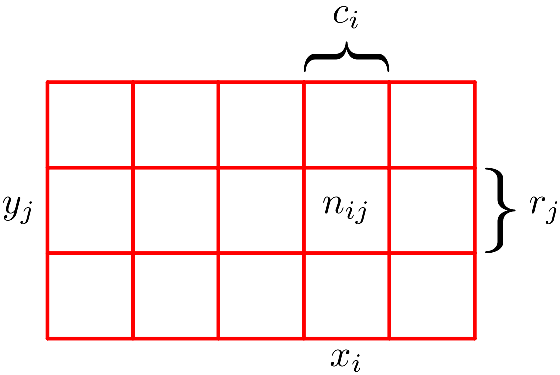
- These slides take you through a few probability basics
- Based on Bishop, Pattern recognition and machine learning, Chapter 1, section 1.2
- Freely available at: `https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/`

Probability



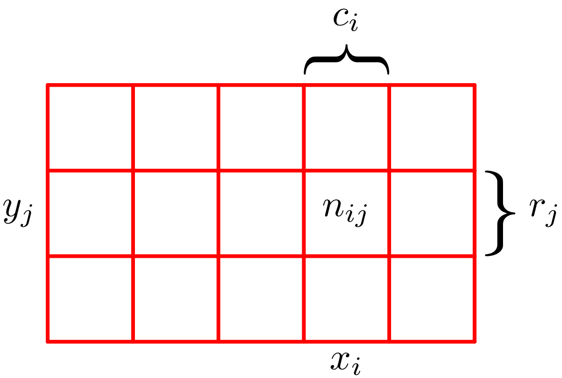
- Suppose the probability of choosing the red box is 0.4, and the probability of choosing the blue box is 0.6.
- We can ask questions such as ‘What is the probability we will get an apple’, or ‘What is the probability that we selected the blue box, given that we have an orange’

Probability



- $P(X = x_i, Y = y_j) = n_{ij}/N$
- $P(X = x_i) = c_i/N$
- $P(Y = y_j | X = x_i) = n_{ij}/c_i$

Sum and product rules



$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

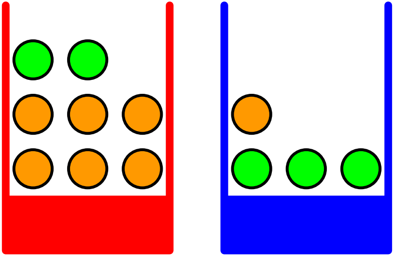
$$P(X = x_i, Y = y_j) = P(Y = y_j | X = x_i) P(X = x_i)$$

Bayes's Theorem

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i | Y = y_j) P(Y = y_j)}{P(X = x_j)}$$

$$P(X = x_i) = \sum_j P(X = x_i | Y = y_j) P(Y = y_j)$$

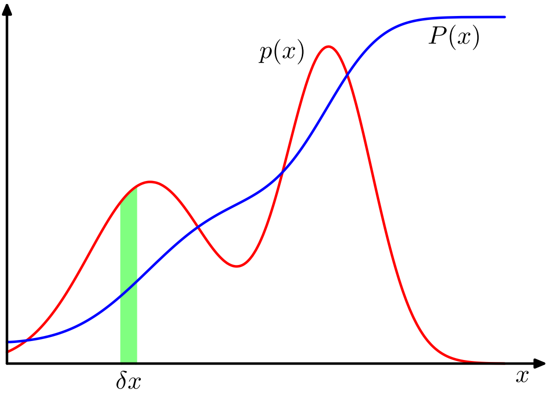
Example



- What is the probability we will get an orange?

$$
\begin{aligned}
P(F = o) &= \sum_j P(F = o|B = y_j)P(B = y_j) \\
&= P(F = o|B = r)P(B = r) + P(F = o|B = b)P(B = b) \\
&= \frac{3}{4} \times \frac{4}{10} + \frac{1}{4} \times \frac{6}{10} = \frac{18}{40} = \frac{9}{20}
\end{aligned}
$$

Probability density



$$p(x \in (a,b)) = \int_a^b p(x)dx$$

$$p(x) = \int p(x,y)dy \qquad p(x,y) = p(y|x)p(x)$$

Expectation and (co-)variance

- The average value of some function $f(x)$ under a probability distribution $p(x)$ is called the expectation of f, $\mathbb{E}[f]$

$$\mathbb{E}[f] = \sum_x p(x)f(x) \tag{1}$$

$$\mathbb{E}[f] = \int p(x)f(x)dx \tag{2}$$

- If we are given a finite number N of points drawn from a probability distribution or density, the expectation can be approximated as a finite sum over these points

$$\mathbb{E}[f] \approx \frac{1}{N}\sum_{n=1}^{N} f(x_n) \tag{3}$$

Expectation and (co-)variance

- The variance of $f(x)$ gives a measure of how much variability there is in $f(x)$ around its mean.

$$
\begin{aligned}
var[f(x)] &= \mathbb{E}(f(x) - \mathbb{E}[f(x)])^2 \\
var(x) &= \mathbb{E}(x - \mathbb{E}[x])^2 \\
&= \mathbb{E}[x^2] - \mathbb{E}[x]^2
\end{aligned}
$$

- The covariance of two random variables x and y expresses the extent to which x and y vary together

$$
\begin{aligned}
cov[x, y] &= \mathbb{E}_{x,y}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] \\
&= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]
\end{aligned}
$$

Covariance matrix

- If $\vec{x}$ and $\vec{y}$ are vectors of random variables, then

$$
\begin{aligned}
cov[\vec{x}, \vec{y}] &= \mathbb{E}_{\vec{x},\vec{y}}[(\vec{x} - \mathbb{E}[\vec{x}])(\vec{y}^T - \mathbb{E}[\vec{y}^T])] \\
&= \mathbb{E}_{\vec{x},\vec{y}}[\vec{x}\vec{y}^T] - \mathbb{E}[\vec{x}]\mathbb{E}[\vec{y}^T]
\end{aligned}
$$

Summary

- We can use probability to calculate how likely it is that an event will occur
- Probability distributions can be discrete or continuous
- We can calculate the probability of an event conditioned on another event, sum over a given variable, or look at the probability of two events happening jointly
- We can calculate the expectation and covariance of a random variable.
- To learn more, take a look at `https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/`