Artificial Intelligence II
Some supplementary notes on probability - part II

Sean B. Holden
March 2010

1 Introduction

These notes provide a reminder of some further simple manipulations that are needed to understand the application of Bayes' theorem to supervised learning. They should be read in conjunction with the earlier Part I of the supplementary notes. Once again, random variables are assumed to be discrete, but all the following results still hold for continuous random variables, with sums replaced by integrals where necessary.

1.1 Some (slightly) unconventional notation

In the machine learning literature there is a common notation intended to make it easy to keep track of which random variables and which distributions are relevant in an expression. While this notation is common within the field, it's rarely if ever seen elsewhere; it is however very useful.

A statistician would define the *expected value* of the random variable X as

$$\mathbb{E}[X] = \sum_{x \in X} x P(x)$$

Here, it is implicit that the probability distribution for X is P. With complex expressions involving combinations of functions defined on random variables with multiple underlying distributions it can be more tricky to keep track of which distributions are relevant. Thus the notation

$$\mathbb{E}_{x \sim P(X)}[f(X)]$$

where f is some function defined on X is intended to indicate explicitly that the distribution of X is P, in situations where we don't write out the full definition

$$\mathbb{E}_{x \sim P(X)}[f(X)] = \sum_{x \in X} f(x) P(x)$$

to make it clear.

1.2 Expected value and conditional expected value

The standard definition of the expected value of a function f of a random variable X is

$$\mathbb{E}_{x \sim P(X)}[f(X)] = \sum_{x \in X} f(x) P(x)$$

as already noted. We can also define the *conditional expected value* of $f(X)$ given Y as

$$\mathbb{E}_{x \sim P(X|Y)}[f(X)] = \sum_{x \in X} f(x) P(x|Y)$$

Now here's an important point: *the value of this expression depends on the value of* Y. Thus, the conditional expected value is itself a function of the random variable Y. What is its expected value? Well

$$
\begin{aligned}
\mathbb{E}_{y \sim P(Y)}\left[\mathbb{E}_{x \sim P(X|Y)}[f(X)]\right] &= \sum_{y \in Y} \mathbb{E}_{x \sim P(X|Y)}[f(X)]\, P(y) \\
&= \sum_{y \in Y} \sum_{x \in X} f(x) P(x|y) P(y) \\
&= \sum_{y \in Y} \sum_{x \in X} f(x) P(x, y) \\
&= \sum_{x \in X} f(x) \sum_{y \in Y} P(x, y) \\
&= \sum_{x \in X} f(x) P(x) \\
&= \mathbb{E}_{x \sim P(X)}[f(X)]
\end{aligned}
$$

or in the more usual notation

$$\mathbb{E}[\mathbb{E}[f(X)|Y]] = \mathbb{E}[f(X)]$$

1.3 Expected value of the indicator function

For any $b \in \{\mathtt{true}, \mathtt{false}\}$ the *indicator function* $\mathbb{I}$ is defined as

$$\mathbb{I}(b) = \begin{cases} 1 & \text{if } b = \mathtt{true} \\ 0 & \text{if } b = \mathtt{false} \end{cases}$$

Let f be a boolean-valued function on a random variable X. Then

$$
\begin{aligned}
\mathbb{E}_{x \sim P(X)}[\mathbb{I}(f(x))] &= \sum_{x \in X} \mathbb{I}(f(x)) P(x) \\
&= \sum_{x \in X, f(x) \text{ is } \mathtt{true}} \mathbb{I}(f(x)) P(x) + \sum_{x \in X, f(x) \text{ is } \mathtt{false}} \mathbb{I}(f(x)) P(x) \\
&= \sum_{x \in X, f(x) \text{ is } \mathtt{true}} P(x) \\
&= \Pr_{x \sim P(x)}[f(x) = \mathtt{true}]
\end{aligned}
$$

This provides a standard method for calculating probabilities by evaluating expected values. So for example if we roll a fair die and consider $f(X)$ to be true if and only if the outcome is even then

$$\Pr(\text{outcome is even}) = \mathbb{E}[\mathbb{I}(f(X))] = 1/6 + 1/6 + 1/6 = 1/2$$

as expected.