Applied Math 9

Handout 1 for Markov Chains: Conditional Probabilities

A standard and intuitive interpretation of probability is in terms "relative frequency." Suppose that each day you record sunny (S), cloudy (C), or raining (R). Over many years you accumulate data, and find that on 4,434 days out of 10,000, it is S. You might then declare the probability that it will be S on a given day to be approximately $0.44 \approx 4,434/10,000$ (here $\approx$ stands for "is approximately equal to"). This interpretation will help in understanding conditional probability. Suppose that we also record the barometric pressure, and keep track of three values: high (H), medium (M), and low (L). Suppose further that the pressure was H for 3,898 days, and that *both* of the outcomes S and H occured 3,576 days. You might be tempted to conclude that there is a connection between the event S and the event H. Indeed, one expects

$$P\{\text{A given day is S and H}\} \approx 0.36.$$

If someone were to tell you that on a particular date the pressure will be H and ask what you think is the probability of S, you would not answer 0.44, since that does not take account of the new information. Instead, you might repeat the logic that lead to 0.44 previously, but *restrict* yourself to only those days that conform to the new data. You might look over your records, and realize that on those days when it was H (3,898 days), the sun was out 3,576 times. The "relative frequency" interpretation of probability would suggest that the probability that a given day is S, *given that the pressure is H*, is $0.92 = 3,576/3,898$. We write this *conditional probability* as

$$P\{\text{A given day is S} \mid \text{It is H}\}.$$

Note that

$$P\{\text{A given day is S} \mid \text{It is H}\} = \frac{P\{\text{A given day is S and H}\}}{P\{\text{A given day is H}\}} = \frac{\frac{3{,}576}{10{,}000}}{\frac{3{,}898}{10{,}000}}.$$

This motivates the following general definition. Let X be a random variable with set of possible outcomes $\{a_i, i = 1, \ldots, n\}$, and let Y be a random variable with outcome space $\{b_j, j = 1, \ldots, m\}$. We define the conditional distribution of X, given that $Y = b_j$, by the formula

$$P\{X = a_i | Y = b_j\} = \frac{P\{X = a_i \text{ and } Y = b_j\}}{P\{Y = b_j\}}.$$

Note that to evaluate the conditional probability, one needs to know the *joint* distribution of the pair X and Y. Thus we need to know all the values

$$p_{(i,j)} = P\{X = a_i \text{ and } Y = b_j\}.$$

Once we have the conditional probabilities, we can define conditional expectations (expected values). For example

$$E[X|Y = b_j] = \sum_{i=1}^{n} a_i P\{X = a_i|Y = b_j\},$$

and

$$E\left[X^2|Y = b_j\right] = \sum_{i=1}^{n} a_i^2 P\{X = a_i|Y = b_j\}.$$

The conditional probabilities represent an improved description of the random quantity X that utilizes our information about Y.

Independence. Now suppose that X and Y are independent. Intuitively, this means that observing Y should tell us nothing about X. In fact, we have

$$\begin{aligned} P\{X = a_i|Y = b_j\} &= \frac{P\{X = a_i \text{ and } Y = b_j\}}{P\{Y = b_j\}} \\ &= \frac{P\{X = a_i\} P\{Y = b_j\}}{P\{Y = b_j\}} \\ &= P\{X = a_i\}, \end{aligned}$$

where the middle equality is due to the definition of independence. Thus if X and Y are independent the conditional probability of X does not depend on the observed value of Y. In a sense, this justifies the definition of independence given in Handout 3 for Games.