Computational Intelligence 2008–2009

Tutorial I-a

Peter Lucas, Marina Velikova and Nivea de Carvalho-Ferreira
Institute for Computing and Information Sciences
University of Nijmegen

Elementary probability theory

We adopt the following notations with respect to probability distributions and Boolean variables. Let X denote a variable; if X is a *binary* variable, e.g. taking either the value *true* or *false*, $X = \mathit{true}$ is also denoted by simply x; similarly, $X = \mathit{false}$ is also referred to by $\neg x$.

Furthermore, when it is not really important to know what value a variable takes, we just refer to the variable itself. For example, $P(X)$ either stands for $P(x)$ or $P(\neg x)$. Of course, a statement like $P(X) = 0.7$ would be silly, as in this case it does matter where X really stands for: if it would be x, then $P(x) = 0.7$ and $P(\neg x) = 0.3$, and if it would be $\neg x$, then $P(\neg x) = 0.7$ and $P(x) = 0.3$.

Finally, the expression $\sum_X P(X)$ means summing over all possible values of the variable X; if X is binary, then

$$\sum_X P(X) = P(x) + P(\neg x)$$

This notation can be generalised for joint probability distributions, e.g.

$$\sum_{X,Y} P(X, Y, a) = P(x, y, a) + P(\neg x, y, a) + P(x, \neg y, a) + P(\neg x, \neg y, a)$$

Of course, the same notation can also be used for conditional probability distributions, as those conform to the axioms of probability theory as well:

$$\sum_X P(X \mid Y) = P(x \mid Y) + P(\neg x \mid Y)$$

Exercise 1

Let P be a probability distribution defined on the set of Boolean expressions $\mathcal{B}$, i.e. the following axioms hold for P:

- $P(\top) = 1$;
- $P(\bot) = 0$;
- $P(x \vee y) = P(x) + P(y)$, if $x \wedge y = \bot$, i.e. x and y are disjoint, $x, y \in \mathcal{B}$.

a. Draw a Venn diagram with x, y and $x \wedge y$, for the case that $x \wedge y \neq \bot$.

b. Now prove that the following holds in general:

$$P(x \vee y) = P(x) + P(y) - P(x \wedge y)$$

Exercise 2

Let P be a *joint* probability distribution, defined as follows (note that $P(A, B)$ is a shorthand for $P(A \wedge B)$).

$P(a, b) = 0.3 \quad P(\neg a, b) = 0.2$
$P(a, \neg b) = 0.4 \quad P(\neg a, \neg b) = 0.1$

a. Compute $P(a)$ and $P(b)$.

b. Compute $P(a \mid b)$.

c. Using these last results, use Bayes' rule to compute $P(b \mid a)$.

Exercise 3

a. Proof that Bayes' rule holds, based on the definition of conditional probabilities:

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

b. Why is this rule sometimes called the *arc reversal* rule?

c. Let $P(B \mid A_1, \ldots, A_n)$ be a conditional probability distribution. Explain why computing $P(B \mid A_1, \ldots, A_n)$ may be computationally hard, and discuss possible solutions.

Exercise 4

Consider joint probability distribution $P(X1, X_2, X_3, X_4)$:

$$\begin{aligned}
P(x_1, x_2, x_3, x_4) &= 0.1 \\
P(x_1, \neg x_2, x_3, x_4) &= 0.04 \\
P(x_1, x_2, \neg x_3, x_4) &= 0.03 \\
P(x_1, x_2, x_3, \neg x_4) &= 0.1 \\
P(\neg x_1, x_2, x_3, x_4) &= 0.0 \\
P(\neg x_1, \neg x_2, x_3, x_4) &= 0.2 \\
P(\neg x_1, x_2, \neg x_3, x_4) &= 0.08 \\
P(\neg x_1, x_2, x_3, \neg x_4) &= 0.1 \\
P(x_1, \neg x_2, \neg x_3, x_4) &= 0.015 \\
P(x_1, \neg x_2, x_3, \neg x_4) &= 0.1 \\
P(x_1, x_2, \neg x_3, \neg x_4) &= 0.004 \\
P(\neg x_1, \neg x_2, \neg x_3, x_4) &= 0.005 \\
P(\neg x_1, \neg x_2, x_3, \neg x_4) &= 0.01
\end{aligned}$$

$$
\begin{aligned}
P(\neg x_1, x_2, \neg x_3, \neg x_4) &= 0.01 \\
P(x_1, \neg x_2, \neg x_3, \neg x_4) &= 0.006 \\
P(\neg x_1, \neg x_2, \neg x_3, \neg x_4) &= 0.2
\end{aligned}
$$

a. Compute $P(x_2 \vee \neg x_3 \mid x_1 \wedge x_4)$.

b. Split up $P(X_1, X_2, X_3, X_4)$ into three factors, and compute the values of these factors for $X_j = \top$, $j = 1, \ldots, 4$.

Exercise 5

Let X, Y and Z three different stochastic variables. Assume that X is conditionally independent of Y given Z.

a. Try to define a probability distribution $P(X \mid Y, Z)$ (i.e. supply the numbers for that distribution) for which the conditional independence property holds.

b. Proof that from $P(X \mid Y, Z) = P(X \mid Z)$ it follows that $P(Y \mid X, Z) = P(Y \mid Z)$.

Exercise 6

a. Prove that if the sets of variables $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given the set of variables $\mathbf{Z}$, i.e.

$$P(\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Z})$$

it follows that

$$P(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Z}) P(\mathbf{Y} \mid \mathbf{Z})$$

b. Explain why the use of marginalisation is often combined with conditioning.

Exercise 7

There are three prisoners, A, B, and C. Two of them will be released and one will be executed. A asks the warden to tell him the name of one of the others in his cohort who will be released. As the question is not directly about A's fate, the warden obliges and says, "B will be released." Assuming the warden's truthfulness, what are A's and C's respective probabilities of dying now? (*Three Prisoners Problem - Martin Gardner*)