Summary of Probability Rules

Tim Bailey

February 28, 2006

This note gives a few key definitions of probability theory and a very brief summary of the three rules for the manipulation of probability distributions.

The *probability density function* (PDF) of an uncertain or *random* variable x is denoted $p(x)$. For an N-dimensional random vector $\mathbf{x}$, the PDF $p(\mathbf{x})$ is defined as a functional mapping $\mathbb{R}^N \mapsto \mathbb{R}$ such that

$$p(\mathbf{x}) \geq 0, \quad \textit{for all } \mathbf{x}$$

and

$$\int p(\mathbf{x})\, d\mathbf{x} = 1$$

The *joint* PDF of vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted $p(\mathbf{x}, \mathbf{y})$. The *marginal* density of $\mathbf{x}$ is the distribution of $\mathbf{x}$ for *any* value of $\mathbf{y}$, and is computed as

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y})\, d\mathbf{y} \tag{1}$$

Two random vectors are *independent* if and only if their joint PDF is equal to the product of their marginal densities.

$$p(\mathbf{x}, \mathbf{y}) \overset{indep}{=} p(\mathbf{x})\, p(\mathbf{y}) \tag{2}$$

The *conditional* density of $\mathbf{x}$ for a particular *given* value of $\mathbf{y}$ is denoted $p(\mathbf{x}|\mathbf{y})$, and defined as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \tag{3}$$

If $\mathbf{x}$ and $\mathbf{y}$ are independent, then $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})$.

Algebraic manipulation of PDFs is based on three rules: the product rule, the sum rule and Bayes theorem. (For generality, we include the value $\mathbf{H}$ in the following definitions, which represents all other known or assumed quantities.) The product rule (or chain rule) is derived from the definition of conditional probability (Equ. 3).

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}|\mathbf{H}) &= p(\mathbf{x}|\mathbf{y}, \mathbf{H})\, p(\mathbf{y}|\mathbf{H}) \\ &= p(\mathbf{y}|\mathbf{x}, \mathbf{H})\, p(\mathbf{x}|\mathbf{H}) \end{aligned} \tag{4}$$

The sum rule follows from the definition of marginalisation (Equ. 1).

$$\begin{aligned} p(\mathbf{x}|\mathbf{H}) &= \int p(\mathbf{x}, \mathbf{y}|\mathbf{H})\, d\mathbf{y} \\ &= \int p(\mathbf{x}|\mathbf{y}, \mathbf{H})\, p(\mathbf{y}|\mathbf{H})\, d\mathbf{y} \end{aligned} \tag{5}$$

The sum rule is also known as the "total probability theorem". Bayes theorem (or Bayes rule) is derived from the product rule.

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}, \mathbf{H}) &= \frac{p(\mathbf{y}|\mathbf{x}, \mathbf{H})\, p(\mathbf{x}|\mathbf{H})}{p(\mathbf{y}|\mathbf{H})} \\ &= \frac{p(\mathbf{y}|\mathbf{x}, \mathbf{H})\, p(\mathbf{x}|\mathbf{H})}{\int p(\mathbf{y}|\mathbf{x}, \mathbf{H})\, p(\mathbf{x}|\mathbf{H})\, d\mathbf{x}} \end{aligned} \tag{6}$$

The denominator of Equ. 6 is independent of $\mathbf{x}$, so we often treat it simply as a normalising constant, such that $p(\mathbf{x}|\mathbf{y}, \mathbf{H})$ integrates to one. For many estimation problems, we are only interested in computing the posterior up to proportionality, and the denominator may be neglected altogether.