Chapter 2
A Primer in Probability

This section provides a summary of probability theory, as necessary to understand this thesis. We consider both discrete and continuous random variables.

▷ A discrete random variable S takes on one of a set of possible values $\mathcal{A}_S = \{a_1, a_2, \ldots, a_I\}$ with probabilities $\mathcal{P}_S = \{p_1, p_2, \ldots, p_I\}$ such that $\text{prob}\,(S = a_i) = p_i,\ p_i \geq 0$ and $\sum_{a_i \in \mathcal{A}_S} \text{prob}\,(S = a_i) = 1$. The probability that S is found in W, a subset of $\mathcal{A}_S$ is

$$\text{prob}\,(S \in W) = \sum_{a_i \in W} \text{prob}\,(S = a_i) \tag{2.1}$$

For a continuous random variable X, we only assign probabilities to ranges of values for X. The probability that $a \leq X \leq b$ is

$$\text{prob}\,(a \leq X \leq b) = \int_a^b \varrho\,(x)\,\mathrm{d}x. \tag{2.2}$$

Here $\varrho\,(x)$ is termed the probability density function (pdf). The pdf is a non-negative, integrable function of x. The differential relation (2.3) is also true provided that the pdf is continuous at x.

$$\frac{\mathrm{d}}{\mathrm{d}x}\text{prob}\,(X \leq x) = \varrho\,(x) \tag{2.3}$$

As in the discrete case, the total probability is one.

$$\int_{-\infty}^{\infty} \varrho\,(x)\,\mathrm{d}x = 1 \tag{2.4}$$

▷ We have the pdf for a random variable X. We desire the pdf for another variable Y which is a single valued function of X: $Y = f\,(X)$. This transformation is called a "change of variables".

R.J.N. Baldock, *Classical Statistical Mechanics with Nested Sampling*,
Springer Theses, https://doi.org/10.1007/978-3-319-66769-0_2

$$\varrho\,(Y = y) = \varrho\,(X = x) \times \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \tag{2.5}$$

$$= \varrho\,(X = x) \div \left|\frac{\mathrm{d}f}{\mathrm{d}x}\right| \tag{2.6}$$

▷ Consider ordered pairs $(S = a_i, T = b_i)$: realisations of the discrete random variables $S \in \mathcal{A}_S = \{a_1, a_2, \ldots, a_I\}$ and $T \in \mathcal{A}_T = \{b_1, b_2, \ldots, b_I\}$. We call the probability distribution over these pairs the joint probability of S and T, $\mathrm{prob}\,(S, T)$.
In a similar manner for two continuous random variables X and Y we define the joint probability distribution via a "multivariate" pdf

$$\mathrm{prob}\,(a \le X \le b, c \le Y \le d) = \int_{x=a}^{b} \int_{y=c}^{d} \varrho\,(X = x, Y = y)\,\mathrm{d}x\mathrm{d}y. \tag{2.7}$$

▷ Two random variables (S and T) or (X and Y) are independent if and only if

$$\mathrm{prob}\,(S, T) = \mathrm{prob}\,(S)\,\mathrm{prob}\,(T) \tag{2.8}$$

$$\varrho\,(X = x, Y = y) = \varrho\,(X = x)\,\varrho\,(Y = y) \tag{2.9}$$

▷ We can recover the probability distribution for S alone $\mathrm{prob}\,(S = a_i) = p_i$ from the joint distribution $\mathrm{prob}\,(S, T)$ by summing over all values of T: a process called marginalisation.

$$\mathrm{prob}\,(S = a_i) = \sum_{b_j \in \mathcal{A}_T} \mathrm{prob}\left(S = a_i, T = b_j\right) \tag{2.10}$$

Likewise, in the continuous case, we can recover the pdf for either variable by integrating the multivariate pdf.

$$\varrho\,(X = x) = \int_{-\infty}^{\infty} \varrho\,(X = x, Y = y)\,\mathrm{d}y \tag{2.11}$$

▷ The conditional probability $\mathrm{prob}\left(S = a_i | T = b_j\right)$ represents "the probability that $S = a_i$ *given* $T = b_j$". It is given by

$$\mathrm{prob}\left(S = a_i | T = b_j\right) \equiv \frac{\mathrm{prob}\left(S = a_i, T = b_j\right)}{\mathrm{prob}\left(T = b_j\right)} \text{ if } \mathrm{prob}\left(T = b_j\right) \ne 0. \tag{2.12}$$

(If $\mathrm{prob}\left(T = b_j\right) = 0$ then $\mathrm{prob}\left(S = a_i | T = b_j\right)$ is undefined.)
If U is the range $[a, b]$ and V the range $[c, d]$

$$\mathrm{prob}\,(X \in U | Y \in V) \equiv \frac{\mathrm{prob}\,(X \in U, Y \in V)}{\mathrm{prob}\,(Y \in V)} \text{ if } \mathrm{prob}\,(Y \in V) \ne 0. \tag{2.13}$$

(Again, if $\mathrm{prob}\,(Y \in V) = 0$ then $\mathrm{prob}\,(X \in U|Y \in V)$ is undefined.)

▷ From the definition of conditional probability we have the "product rule" (also called the "chain rule").

$$\begin{aligned}\mathrm{prob}\left(S = a_i, T = b_j\right) &= \mathrm{prob}\left(S = a_i|T = b_j\right) \times \mathrm{prob}\left(T = b_j\right)\\ &= \mathrm{prob}\left(T = b_j|S = a_i\right) \times \mathrm{prob}\,(S = a_i)\end{aligned} \tag{2.14}$$

$$\begin{aligned}\varrho\,(X = x, Y = y) &= \varrho\,(X = x|Y = y) \times \varrho\,(Y = y)\\ &= \varrho\,(Y = y|X = x) \times \varrho\,(X = x)\end{aligned} \tag{2.15}$$

▷ The sum rule is obtained by rewriting marginalisation using the product rule.

$$\begin{aligned}\mathrm{prob}\,(S = a_i) &= \sum_{b_j \in \mathcal{A}_T} \mathrm{prob}\left(S = a_i, T = b_j\right)\\ &= \sum_{b_j \in \mathcal{A}_T} \mathrm{prob}\left(S = a_i|T = b_j\right) \times \mathrm{prob}\left(T = b_j\right)\end{aligned} \tag{2.16}$$

$$\begin{aligned}\varrho\,(X = x) &= \int_{-\infty}^{\infty} \varrho\,(X = x, Y = y)\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty} \varrho\,(X = x|Y = y) \times \varrho\,(Y = y)\,\mathrm{d}y\end{aligned} \tag{2.17}$$

▷ We obtain Bayes' theorem by rearranging the product rule.

$$\mathrm{prob}\,(S|T) = \frac{\mathrm{prob}\,(T|S) \times \mathrm{prob}\,(S)}{\mathrm{prob}\,(T)} \tag{2.18}$$

$$\mathrm{prob}\,(S|T) = \frac{\mathrm{prob}\,(T|S) \times \mathrm{prob}\,(S)}{\sum_S \mathrm{prob}\,(T|S) \times \mathrm{prob}\,(S)} \tag{2.19}$$

$$\mathrm{prob}\,(X \in U|Y \in V) = \frac{\mathrm{prob}\,(Y \in V|X \in U) \times \mathrm{prob}\,(X \in U)}{\mathrm{prob}\,(Y \in V)} \tag{2.20}$$

$$\mathrm{prob}\,(X \in U|Y \in V) = \frac{\mathrm{prob}\,(Y \in V|X \in U) \times \mathrm{prob}\,(X \in U)}{\int_{y \in V} \int_{x=-\infty}^{\infty} \varrho\,(Y = y|X = x) \times \varrho\,(X = x)\,\mathrm{d}x\mathrm{d}y} \tag{2.21}$$

http://www.springer.com/978-3-319-66768-3

Classical Statistical Mechanics with Nested Sampling
Baldock, R.J.N.
2017, XII, 144 p. 30 illus., 25 illus. in color., Hardcover
ISBN: 978-3-319-66768-3