Chapter 4

Probability

4.1 Random Variable

4.1.1 Describing random change

While motion in the realm of macroscopic bodies is largely deterministic, in the realm of microscopic motion, stochasticity is the rule. A prominent classical example is provided by motion of colloids in a viscous fluid. A large stone drops in such a fluid falls with a uniform (terminal) velocity set by the balance of the force of gravity and fluid friction. A micron-sized particle in the fluid, however, if observed with a microscope, performs a jittery motion that only on average moves in the direction of gravity. The Scottish botanist, Robert Brown, first discussed these fluctuations in 1827. To discuss and analyze such *Brownian motion* a new mathematical perspective is needed, based on the concept of probability, which is the topic of the next part of this material.

The simplest example of a random variable is a coin toss that can come up head or tails. More generally, a random variable X may have a set of possible outcomes $\mathcal{S} \in \{x_1, x_2, \cdots, x_n\}$, e.g. $n = 6$ for outcomes of throwing a dice. To various outcomes of the random variable, we then assign probabilities, which must satisfy the following conditions:

- *Positivity*: $p_i \geq 0$, i.e. all probabilities must be real and non-negative.

- *Additivity*: Probabilities of independent outcomes is additive, e.g. the probability of an even number in the throw of a dice is the sum of probabilities for obtaining 2, 4, and 6.

- *Normalization*: $p(\mathcal{S}) = \infty$, i.e. the random variable must have take one of the possible set of outcomes.

In principle, there are two approaches to assigning probabilities:

- *Objective* probabilities are obtained *experimentally* from the relative frequency of the occurrence of an outcome in many tests of the random variable. If the random process

is repeated N times, and the event A occurs N_A times, then

$$p(A) = \lim_{N \to \infty} \frac{N_A}{N}.$$

For example, a series of $N = 100$, 200, 300 throws of a dice may result in $N_1 = 19$, 30, 48 occurrences of 1. The ratios .19, .15, .16 provide an increasingly more reliable estimate of the probability $p_{\text{dice}}(\{1\})$.

- *Subjective* probabilities provide a *theoretical* estimate based on the uncertainties related to lack of precise knowledge of outcomes. For example, the assessment $p_{\text{dice}}(\{1\}) = 1/6$, is based on the knowledge that there are six possible outcomes to a dice throw, and that in the absence of any prior reason to believe that the dice is biased, all six are equally likely. The consequences of such subjective assignments of probability have to be checked against measurements, and they may need to be modified as more information about the outcomes becomes available.

4.1.2 Moments and cumulants

Quite generally, the *expectation value* of any function $F(X)$ of the random variable X with outcomes $\{x_i\}$ is given by

$$\langle F(X) \rangle = \sum_i p_i F(x_i). \tag{4.1.1}$$

- Of particular relevance are the *mean* or *average* of X, obtained as

$$\langle X \rangle = \sum_i p_i x_i. \tag{4.1.2}$$

Similarly, higher *moments* of the random variable are expectation values for powers of the random variable; the ℓth moment given by

$$\langle X^\ell \rangle = \sum_i p_i x_i^\ell. \tag{4.1.3}$$

- A quite useful method for obtaining moments of a probability distribution function is to employ a so called *moment generating function*, which is

$$G(\lambda) = \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} \langle X^\ell \rangle = \langle e^{\lambda X} \rangle. \tag{4.1.4}$$

Moments of the random variable can then be generated as terms of the coefficients of the Taylor expansion of $G(\lambda)$ around the origin.

- Another useful quantity is the *cumulant generating function* which is the logarithm of the moment characteristic function. Its expansion around the origin generates the *cumulants* of the random variable, defined through

$$\ln G(\lambda) = \sum_{\ell=1}^{\infty} \frac{\lambda^{\ell}}{\ell!} \langle X^{\ell} \rangle_c . \tag{4.1.5}$$

- The first four cumulants are called the *mean, variance, skewness,* and *curtosis* (or kurtosis) of the random variable respectively, and are obtained from the moments as

$$\langle X \rangle_c = \langle X \rangle ,$$
$$\langle X^2 \rangle_c = \langle X^2 \rangle - \langle X \rangle^2 ,$$
$$\langle X^3 \rangle_c = \langle X^3 \rangle - 3 \langle X^2 \rangle \langle X \rangle + 2 \langle X \rangle^3 ,$$
$$\langle X^4 \rangle_c = \langle X^4 \rangle - 4 \langle X^3 \rangle \langle X \rangle - 3 \langle X^2 \rangle^2 + 12 \langle X^2 \rangle \langle X \rangle^2 - 6 \langle X \rangle^4 . \tag{4.1.6}$$

- An important theorem allows easy computation of moments in terms of the cumulants: Represent the nth cumulant graphically as a *connected cluster* of n points. The mth moment is then obtained by summing all possible subdivisions of m points into groupings of smaller (connected or disconnected) clusters. The contribution of each subdivision to the sum is the product of the connected cumulants that it represents. Using this result the first four moments are computed by adding configurations as

$$\langle X \rangle = \langle X \rangle_c ,$$
$$\langle X^2 \rangle = \langle X^2 \rangle_c + \langle X \rangle_c^2 ,$$
$$\langle X^3 \rangle = \langle X^3 \rangle_c + 3 \langle X^2 \rangle_c \langle X \rangle_c + \langle X \rangle_c^3 ,$$
$$\langle X^4 \rangle = \langle X^4 \rangle_c + 4 \langle X^3 \rangle_c \langle X \rangle_c + 3 \langle X^2 \rangle_c^2 + 6 \langle X^2 \rangle_c \langle X \rangle_c^2 + \langle X \rangle_c^4 . \tag{4.1.7}$$

4.1.3 Bionomial distribution

Consider a random variable with two outcomes A and B (e.g. a coin toss) of relative probabilities p_A and $p_B = 1 - p_A$. The probability that in N trials the event A occurs exactly N_A times (e.g. 5 heads in 12 coin tosses), is given by the binomial distribution

$$p_N(N_A) = \left(\frac{N!}{N_A!(N-N_A)!} \right) p_A^{N_A} p_B^{N-N_A} . \tag{4.1.8}$$

The prefactor is just the coefficient obtained in the binomial expansion of $(p_A + p_B)^N$, and gives the number of possible orderings of N_A events A and $N_B = N - N_A$ events B.

The generating function for the binomial distribution is

$$G_N(\lambda) = \langle e^{\lambda N_A} \rangle = \sum_{N_A=0}^{N} \frac{N!}{N_A!(N-N_A)!} p_A^{N_A} p_B^{N-N_A} e^{\lambda N_A} = \left(p_A e^{\lambda} + p_B \right)^N . \tag{4.1.9}$$

The resulting cumulant generating function is

$$\ln G_N(\lambda) = N \ln \left(p_A e^{-ik} + p_B \right) = N \ln G_1(\lambda), \tag{4.1.10}$$

where $\ln G_1(\lambda)$ is the cumulant generating function for a single step. Hence, the cumulants after N steps are simply N times the cumulants in a single step. In each step, the allowed values of N_A are 0 and 1 with respective probabilities p_B and p_A, leading to $\langle N_A^\ell \rangle = p_A$, for all ℓ. After N trials the first two cumulants are

$$\langle N_A \rangle_c = N p_A \quad , \quad \langle N_A^2 \rangle_c = N \left(p_A - p_A^2 \right) = N p_A p_B \quad . \tag{4.1.11}$$

A measure of fluctuations around the mean is provided by the *standard deviation*, which is the square root of the variance. While the mean of the binomial distribution scales as N, its standard deviation only grows as $\sqrt{N}$. Hence, the *relative* uncertainty becomes smaller for large N.

The binomial distribution is straightforwardly generalized to a *multinomial* distribution, when the several outcomes $\{A, B, \cdots, M\}$ occur with probabilities $\{p_A, p_B, \cdots, p_M\}$. The probability of finding outcomes $\{N_A, N_B, \cdots, N_M\}$ in a total of $N = N_A + N_B \cdots + N_M$ trials is

$$p_N\left(\{N_A, N_B, \cdots, N_M\}\right) = \frac{N!}{N_A! N_B! \cdots N_M!} p_A^{N_A} p_B^{N_B} \cdots p_M^{N_M} \quad . \tag{4.1.12}$$

4.1.4   Poisson distribution

The classical example of a Poisson process is radioactive decay. Observing a piece of radioactive material over a time interval T shows that:

- The probability of one and only one event (decay) in the interval $[t, t+dt]$ is proportional to dt as $dt \to 0$,

- The probabilities of events at different intervals are independent of each other.

The probability of observing exactly M decays in the interval T is given by the Poisson distribution. It is obtained as a limit of the binomial distribution by subdividing the interval into $N = T/dt \gg 1$ segments of size dt. In each segment, an event occurs with probability $p = \alpha dt$, and there is no event with probability $q = 1 - \alpha dt$. As the probability of more than one event in dt is too small to consider, the process is equivalent to a binomial one. Using Eq. (4.1.10) the generating function for this process is obtained as

$$G(\lambda) = \left(p e^\lambda + q \right)^n = \lim_{dt \to 0} \left[1 + \alpha dt \left(e^\lambda - 1 \right) \right]^{T/dt} = \exp\left[\alpha (e^\lambda - 1) T \right] \quad . \tag{4.1.13}$$

The cumulants of the distribution are obtained from the expansion

$$\ln G(\lambda) = \alpha T (e^\lambda - 1) = \alpha T \sum_{n=1}^{\infty} \frac{(\lambda)^n}{n!}, \quad \implies \quad \langle M^n \rangle_c = \alpha T \quad . \tag{4.1.14}$$

All cumulants have the same value, and the moments are obtained as

$$\langle M \rangle = (\alpha T), \quad \langle M^2 \rangle = (\alpha T)^2 + (\alpha T), \quad \langle M^3 \rangle = (\alpha T)^3 + 3(\alpha T)^2 + (\alpha T). \qquad (4.1.15)$$

Using a similar limiting procedure on the binomial distribution leads to

$$p_{\alpha T}(M) = e^{-\alpha T} \frac{(\alpha T)^M}{M!}. \qquad (4.1.16)$$

Example: Assuming that stars are randomly distributed in the galaxy (clearly unjustified) with a density n, what is the probability that the nearest star is at a distance R?

Since, the probability of finding a star in a small volume dV is ndV, and they are assumed to be independent, the number of stars in a volume V is described by a Poisson process as in Eq. (4.1.16), with $\alpha = n$. The probability $p(R)$, of encountering the first star at a distance R is the product of the probabilities $p_{nV}(0)$, of finding zero stars in the volume $V = 4\pi R^3/3$ around the origin, and $p_{ndV}(1)$, of finding one star in the shell of volume $dV = 4\pi R^2 dR$ at a distance R. Both $p_{nV}(0)$ and $p_{ndV}(1)$ can be calculated from Eq. (4.1.16), and

$$p(R)dR = p_{nV}(0)\, p_{ndV}(1) = e^{-4\pi R^3 n/3}\, e^{-4\pi R^2 ndR}\, 4\pi R^2 ndR,$$

$$\implies \quad p(R) = 4\pi R^2 n \exp\left(-\frac{4\pi}{3} R^3 n\right). \qquad (4.1.17)$$