NPTEL COURSE
PHYSICAL APPLICATIONS OF STOCHASTIC PROCESSES

V. Balakrishnan
Department of Physics, Indian Institute of Technology Madras
Chennai 600 036, India

This course comprises 29 lectures. The notes that follow only deal with the topics discussed in Lectures 1 to 11. Exercises are marked with a star. The extended 'Quiz' following the notes covers both this material as well as the topics discussed in the rest of the course (Lectures 12 to 29). Answers and solutions are given at the end of this write-up. The impertinence of Autocorrect is responsible for any typos that remain!

Reference books and suggested further reading

Balakrishnan V: Elements of Nonequilibrium Statistical Mechanics (Ane Books & CRC Press).
Beck C and Schlögl F: Thermodynamics of Chaotic Systems (Cambridge University Press).
Berg H C: Random Walks in Biology (Princeton University Press).
Cox D R and Miller H D: The Theory of Stochastic Processes (Chapman & Hall).
Denker M and Woyczynski W A: Introductory Statistics and Random Phenomena (Birkhauser).
Doi M and Edwards S F: The Theory of Polymer Dynamics (Cambridge University Press).
Doyle P G and Snell J L: Random Walks and Electrical Networks (Mathematical Association of America).
Feller W: An Introduction to Probability Theory and Its Applications, Vols, 1 & 2 (Wiley).
Gardiner C W: Handbook of Stochastic Processes (Springer).
Grimmett G and Stirzaker D: Probability and Random Processes (Oxford University Press).
Kac M: Probability and Related Topics in Physical Sciences (Wiley-Interscience).
Papoulis A: Probability, Random Variables and Stochastic Processes (McGraw-Hill).
Risken H: The Fokker-Planck Equation: Methods of Solution and Applications (Springer).
Stratonovich R L : Topics in the Theory of Random Noise, Vols. 1 and 2 (Gordon & Breach).
Van Kampen N G: Stochastic Processes in Physics and Chemistry (North-Holland).
Wax N: Selected Papers in Noise and Stochastic Processes (Dover).
Weiss G H: Aspects and Applications of the Random Walk (North-Holland).
Wong E: Introduction to Random Processes (Springer).

Table of contents

distribution.

Lecture 9: One-dimensional random walk. Cumulants of the displacement. Long-time behaviour. Biased random walk on a d-dimensional lattice. Explicit solution, asymptotic behavior. General birth-and-death process with one-step transition probabilities. Master equation. Poisson process as a birth process. Non-stationary Poisson process, solution. Case of linear transition probabilities.

Lecture 10: Stationary distribution for a one-step birth-and-death process. Applications: Sedimentation and the barometric distribution; a chemical reaction; radiation; a population model. Closed set of equations for the mean and variance in a birth-and-death process with linear coefficients.

Lecture 11: Continuous Markov processes. Chapman-Kolmogorov equation. Transition rate and the master equation. Kramers-Moyal expansion. Pawula's Theorem. Diffusion processes. Fokker-Planck equation (FPE). Drift and diffusion terms. Ornstein-Uhlenbeck process. Continuum limit of a biased random walk on a linear lattice: diffusion in a constant force field. Smoluchowski equation. Stationary solution. Stochastic differential equation (SDE) for a diffusion process. Wiener process.

Lecture 12: Correspondence between Langevin and Fokker-Planck equations. Langevin equation for the velocity of a diffusing particle. Fluctuation-dissipation relation. FPE for the conditional density of the velocity. Conditional mean and variance of the velocity. Gaussian solution of the FPE. Velocity autocorrelation function.

Lecture 13: Ornstein-Uhlenbeck distribution. Doob's Theorem. Mean squared displacement. Relation between the diffusion constant and the viscosity. General Kubo-Green formula for the diffusion constant. FPE for the joint density in position and velocity. Long-time limit of the joint density.

Lecture 14: Coupled Langevin equations for the position and velocity. Fokker-Planck equation for the joint density. Langevin and Fokker-Planck equations for a harmonically bound particle. Stationary joint PDF. Relaxation times for position and velocity. High friction limit and corresponding FPE. Particle in a potential. General form of the SDE–FPE correspondence. FPE for the joint position and velocity for a particle in an arbitrary potential: Kramers equation. Preliminary remarks on diffusion of a charged particle in a magnetic field.

Lecture 15: Langevin equation for the velocity of a particle moving in three dimensions. Velocity autocorrelation function for a free particle and in the presence of a magnetic field. Mathematical Brownian motion: SDE and FPE (the

diffusion equation). Autocorrelation function. Non-stationarity of the process. Properties of Brownian trails. Non-differentiability and Hölder continuity. Lévy's arcsine law. Khinchin's law of the iterated logarithm. Relation between a general Gaussian Markov process and Brownian motion. Relation between the Ornstein-Uhlenbeck process and Brownian motion.

Lecture 16: Itô and Fokker-Planck equations for a diffusion process. Diffusion equation in d dimensions. PDF of the radial distance. Its FPE and Itô equation. The FPE and Itô equation for the square of the radial distance (Bessel process). PDF, FPE and Itô equations for the square, n^{th} power and exponential of Brownian motion. Geometric Brownian motion and the Black-Scholes equation.

Lecture 17: Level-crossing statistics of a continuous random process. Mean number and mean rate of threshold crossings. Up-crossings and down-crossings. Example of the Brownian oscillator. Distribution of the extrema of a random process. Preliminary remarks on the power spectral density of a random process. Examples: white noise, exponentially correlated noise.

Lecture 18: Diffusion of a charged particle in a magnetic field. Langevin and Fokker-Planck equations. Transformation to a rotating frame of reference. Exact solution of the FPE for the PDF of the velocity. Correlation matrix of the velocity. Evaluation of the diffusion tensor from the Kubo-Green formula. Transverse and longitudinal diffusion coefficients. Diffusion limit in the presence of a magnetic field.

Lecture 19: Power spectrum of noise: definition. Expression as the Fourier transform of the autocorrelation: Wiener-Khinchin Theorem. Relation between input and output power spectra in an SDE. Transfer function. Multi-component case. Hermiticity of the power spectrum matrix. Thermal noise in a resistor.

Lecture 20: Elements of linear response theory. Example of a diffusing particle under an applied time-dependent force. Dynamic mobility. Relation between the power spectrum of spontaneous fluctuations and the generalized susceptibility: the first fluctuation-dissipation theorem. Second fluctuation-dissipation theorem. Causal, linear, retarded response. Generalised susceptibility as the Fourier-Laplace transform of the response function. Classical formula for the response function. Symmetry and analyticity properties of the frequency-dependent susceptibility. Example: complex admittance of a series LCR circuit.

Lecture 21: Random pulse sequences (generalized shot noise). Poisson pulse process. Covariance of a Poisson pulse process. Mean and variance of a general pulse process: Campbell's Theorem. Extension to higher order cumulants. Power spectrum of generalized shot noise: Carson's Theorem. Power spectrum

of rectangular pulses. Barkhausen noise. The integral of a dichotomous Markov process: dichotomous diffusion. Master equations for the positional PDFs.

Lecture 22: Covariance function of the Poisson process. Discrete analog of dichotomous diffusion: persistent random walk on a linear lattice. Passage to the continuum limit: master equations for dichotomous diffusion on a line. Initial conditions. Exact solution for the PDF of the position. Mean squared displacement. Asymptotic behaviour of the solution in the diffusion limit.

Lecture 23: Level-crossing statistics in dichotomous diffusion. Mean and variance of the number of threshold crossings in a time interval. First-passage time (FPT) distribution for ordinary diffusion. Mean FPT for ordinary diffusion and dichotomous diffusion. FPT distribution for a random walk on a linear lattice. Random walk dimension. Backward Kolmogorov equation for the distribution and moments of the FPT on a general lattice.

Lecture 24: Renewal equation for the first-passage time distribution. Laplace transform of the FPT distribution and its properties. Examples: diffusion on a line and random walk on a linear lattice. FPT distribution from the backward Kolmogorov equation for a general diffusion process. Random walk dimension. Fractals, fractal dimension.

Lecture 25: Illustration of a non-standard random walk dimension: random walk on the Sierpinski gasket. Recurrence and first passage in a Markov chain: formulas for the generating functions of FPT and recurrence time distributions. Criterion for the probability of recurrence. Probability of return to the origin in biased and unbiased random walks on a linear lattice. Divergence of the mean FPT and mean recurrence time in an unbiased random walk.

Lecture 26: Return to the origin in an unbiased random walk on a lattice in d-dimensions. Recurrence for $d \leq 2$ and transience for $d > 2$. Continuous time random walks. Exponential waiting-time density and Markovian walks. Non-exponential waiting-time density and non-Markovian walks.

Lecture 27: Biased continuous time random walk on a lattice in d dimensions. Formal exact solution for the generating function and positional probability distribution. Asymptotic (long-time) behaviour of the mean displacement and its variance for a general non-Markovian random walk. Criterion for sub-diffusive behaviour: anomalous diffusion.

Lecture 28: Probabilistic aspects of deterministic dynamical systems. Phase trajectories in an autonomous dynamical system. Discrete-time dynamics: maps. Coarse-graining of phase space. Invariant measure. Frobenius-Perron equation.

Equivalence of time average and ensemble average: ergodicity. Definitions of escape and recurrence time distributions. Statement of Poincaré recurrence theorem.

Lecture 29: Sojourn probability distribution. Relation between sojourn and escape time distributions. Recurrence time distribution as the second difference of sojourn time distributions. Proof of Poincaré recurrence theorem. Smoluchowski's modified expression for the recurrence time distribution. Illustration of recurrence time statistics for the coarse-grained dynamics of the Bernoulli shift map.

1 Discrete probability distributions

1.1 Mean and variance

Let's begin with a quick recapitulation of some elementary properties of a random variable n that can take on values in the set of integers (or some subset of integers), called the **sample space** of n. Let P_n be the normalized probability that the variable takes on the particular value n. Thus $P_n \geq 0$, and $\sum_n P_n = 1$. The mean value, mean squared value, and variance of n are given by

$$\langle n \rangle \stackrel{\text{def.}}{=} \sum_n n\, P_n\,, \;\; \langle n^2 \rangle \stackrel{\text{def.}}{=} \sum_n n^2\, P_n\,, \;\; \text{Var } n \stackrel{\text{def.}}{=} \langle n^2 \rangle - \langle n \rangle^2 = \langle \big(n - \langle n \rangle\big)^2 \rangle,$$

respectively. The final equation above shows that the variance is strictly positive for a random variable. It vanishes if and only if n takes on just one value with probability 1, i.e., if n is a *sure* or *deterministic* variable rather than a random variable. The standard deviation (often denoted by Δn) is the square root of the variance. It is the basic measure of the *scatter* of the random variable about its mean value. In dimensionless units, one often uses the ratio

$$\text{relative fluctuation } = \Delta n / \langle n \rangle$$

to measure this scatter about the mean. More information about the probability distribution P_n is contained in the higher moments $\langle n^k \rangle$, where $k \geq 3$.

★ **1.** A pair of (distinguishable) dice is tossed once. Each die can give a score of 1, 2, 3, 4, 5, or 6. Let s denote the total score of the pair of dice. It is obvious that the sample space of s is the set of integers from 2 to 12.

(a) Write down the set of probabilities $\{P_s\}$. What is the most probable value of s?

(b) Find the mean, standard deviation and relative fluctuation of s.

Remark: It is useful to recall the language of statistical mechanics. Each possible configuration of the two dice is an **accessible microstate** of the system (which comprises the two dice). The dice are distinguishable, which is why we must count the configuration $(5, 6)$ as different from the configuration $(6, 5)$, and so on. Hence there are $6 \times 6 = 36$ possible microstates, each with an *a priori* (or before-the-event) probability of $\frac{1}{36}$. The total score s may be taken to label the possible **macrostates** of the system. There are 11 possible macrostates, corresponding to $s = 2, 3, \ldots, 12$. The most probable score (i.e., the most probable macrostate) is $s = 7$, because it has the largest number $(= 6)$ of accessible microstates. This is why $P_7 = 6 \times \frac{1}{36} = \frac{1}{6}$.

★ **2.** The probability of a score $s = 3$ is of course $P_3 = 2 \times \frac{1}{36} = \frac{1}{18}$.

(a) Suppose the two dice were made indistinguishable from each other for all practical purposes. What would P_3 be in this case?

(b) Here is a quantum mechanical analog of a pair of dice: Consider a pair of identical bosons that obey quantum mechanics. Suppose the possible values of the energy of each particle are $1, 2, 3, 4, 5$ and 6 (in suitable units), with a corresponding normalized eigenstate $|j\rangle$, where j runs from 1 to 6. Let $|jk\rangle$ denote the normalized state of the two-particle system in which particle 1 has energy j and particle 2 has energy k. What are the possible normalized energy eigenstates of the system? If, further, the system can be in any of these states with equal probability, what is the probability that its energy is equal to 3?

(c) Now consider a pair of identical fermions instead of bosons, all the other conditions remaining the same as in the foregoing. What is the probability that the total energy of the pair is equal to 3?

1.2 Bernoulli trials and the binomial distribution

Consider an event in which the outcome can be either success, with an *a priori* (or before-the-event) probability p; or failure, with an *a priori* probability $q \equiv 1-p$. Here $0 < p < 1$ in general. Such an event is called a **Bernoulli trial**. A simple example is a coin toss, in which the outcome can be a head (success, say) or a tail (failure). If the coin is a *fair* one, then $p = q = \frac{1}{2}$; if the coin is *biased*, then $p \neq \frac{1}{2}$. We make N trials. The question is: what is the probability P_n that the total number of successes is n? Here n is the random variable, and the sample space is the set of integers from 0 to N.

The trials are independent of each other. Hence the probability of any particular sequence of successes (heads) and failures (tails), such that the total number of successes is n, is just $p^n q^{N-n}$. Each such sequence contributes to P_n, and there are ${}^N C_n$ such sequences. Therefore

$$P_n = \binom{N}{n} p^n \, q^{N-n}, \quad 0 \leq n \leq N.$$

This probability distribution is called the **binomial distribution** for an obvious reason: P_n is a term in the binomial expansion of $(p+q)^N$. The binomial distribution is a two-parameter distribution, specified by the parameters p and N.

We are interested in various moments of P_n. These are most easily obtained from the generating function of P_n. This function is given, for any fixed value of N, by

$$f(z) \stackrel{\text{def.}}{=} \sum_{n=0}^{N} P_n \, z^n = (pz+q)^N,$$

where z is a complex variable. Generating functions are usually power series in some variable. When this is an infinite series, the question of its convergence arises. Recall that a series in non-negative integer powers of a complex variable z is absolutely convergent inside some circle of convergence centered at the origin in the complex z-plane. In the present instance, the generating function is just a polynomial of order N in z, which therefore remains a valid representation of the generating function for all finite $|z|$.

The normalization of the probability distribution is trivially verified, since $\sum_{n=0}^{N} P_n = f(1) = 1$. The k^{th} **factorial moment** of n is given by

$$\langle n(n-1)\,\ldots\,(n-k+1)\rangle = \left.\frac{d^k f(z)}{dz^k}\right|_{z=1} .$$

★ **3.** Verify that the mean and variance of the binomial distribution are given by

$$\langle n\rangle = N\,p \quad \text{and} \quad \text{Var}\; n = N\,p\,q,$$

respectively, and that its k^{th} factorial moment is

$$\langle n(n-1)\,\ldots\,(n-k+1)\rangle = \begin{cases} N(N-1)\,\ldots\,(N-k+1)\,p^k\,, & k \le N \\ 0\,, & k > N. \end{cases}$$

★ **4.** What is the condition that must be satisfied in order that the most probable value of n is (a) 0 (b) N?

It is often more convenient to consider another, related generating function, the **moment generating function**. If n is the random variable and u denotes a complex variable, this is defined as

$$M(u) = \langle e^{un}\rangle \equiv \sum_n P_n\, e^{un} = \sum_n P_n\, (e^u)^n.$$

Therefore $M(u)$ is related to the generating function $f(z)$ by

$$M(u) = f(e^u).$$

The advantage of using $M(u)$ rather than $f(z)$ is that the derivatives of $M(u)$ at $u = 0$ directly give the moments (rather than the factorial moments) of the random variable n: we have

$$M(u) = \sum_{k=0}^{\infty} \langle n^k\rangle\, \frac{u^k}{k!} \quad \text{so that} \quad \langle n^k\rangle = \left.\frac{d^k M}{du^k}\right|_{u=0} .$$

The moment generating function of the binomial distribution is

$$M(u) = (pe^u + 1 - p)^N = (pe^u + 1 - p)^{\mu/p},$$

where the final expression follows upon eliminating N in favor of the mean value $\mu = Np$.

1.3 Number fluctuations in a classical ideal gas

The relative fluctuation in the case of the binomial distribution is

$$\frac{\Delta n}{\langle n \rangle} = \sqrt{\frac{q}{Np}} = \sqrt{\frac{q}{\langle n \rangle}}\,.$$

Note the $1/\sqrt{N}$ fall-off of this quantity with increasing N. This feature is not restricted to the binomial distribution. It is characteristic of numerous physical situations in equilibrium statistical mechanics. In that context, N is usually proportional to the number of degrees of freedom of the system concerned. This is essentially why thermodynamics, which deals only with *mean* values of macroscopic observables, and neglects *fluctuations* in them, provides a satisfactory description of physical phenomena under the conditions in which it is applicable. If N is of the order of Avogadro's number, for instance, then $N^{-1/2} \sim 10^{-12} \ll 1$. The following simple physical example illustrates the point.

Consider a (classical, ideal) gas of N molecules in thermal equilibrium in a container of volume V. The molecules move about at random, suffering elastic collisions with each other. The average number density of the particles is $\rho = N/V$. Now consider a sub-volume v in the container. It is obvious that the number of molecules that are present in the sub-volume is a rapidly fluctuating quantity. In principle, this number can be any integer from 0 to N at any given instant of time. Let n be this random number (at a any given instant of time). We want the probability P_n that there are exactly n molecules in the sub-volume. The *a priori* probability that a given molecule is in the sub-volume is just v/V, because it has an equal probability of being in any volume element in the container. Therefore the probability that it lies outside v is $(1 - v/V)$. (The molecules are assumed to be independent of each other, i.e., there is no correlation between the locations of different molecules.) Since the n molecules inside V can be chosen in ${}^N C_n$ ways, we have

$$P_n = \binom{N}{n} \left(\frac{v}{V}\right)^n \left(1 - \frac{v}{V}\right)^{N-n} = \binom{N}{n} \left(\frac{\rho v}{N}\right)^n \left(1 - \frac{\rho v}{N}\right)^{N-n}.$$

Thus P_n is a binomal distribution, with $p = v/V = \rho v/N$. The mean value is given by $\langle n \rangle = \rho v$, as expected. The relative fluctuation in n is $[(1-\rho v/N)/\rho v]^{1/2}$.

★ **5.** The so-called **thermodynamic limit** of this system corresponds to letting $N \to \infty$ and $V \to \infty$, keeping their ratio $N/V = \rho$ fixed at a finite value. In this limit, the binomial distribution goes over into the Poisson distribution, as we'll see. The sample space of n is now the set of all nonnegative integers. The mean value $\langle n \rangle$ remains equal to ρv, of course. The relative fluctuation becomes $(\rho v)^{-1/2}$. Estimate this quantity for a volume $v = 1\,\mathrm{m}^3$ of air at standard temperature and pressure.

1.4 The geometric distribution

Consider once again a coin that has *a priori* probabilities p for heads and $q \equiv 1-p$ for tails, each time it is tossed. The experiment now consists of tossing the coin repeatedly till heads is obtained *for the first time*. What is the probability P_n that this happens on the $(n+1)^{\text{th}}$ toss?

It is clear that the random variable in this case is n. Its sample space is the set of non-negative integers 0, 1, ... *ad inf.* The probability distribution of n is obviously given by

$$P_n = q^n\, p.$$

Such a probability distribution is called a **geometric distribution**, because the successive terms in the sequence $P_0\,,\ P_1\,,\ \ldots$ form a geometric progression. The properties of a geometric distribution are easily deduced.

★ **6.** Show that the generating function for P_n is $f(z) = p/(1-qz)$. Hence show that $\langle n \rangle = q/p\,,\ \mathrm{Var}\,(n) = q/p^2\,,$ and $\Delta n/\langle n \rangle = 1/\sqrt{q}$.

Denoting the mean $\langle n \rangle$ by μ, the geometric distribution and its generating function are given by

$$\boxed{P_n = \frac{1}{\mu+1}\left(\frac{\mu}{\mu+1}\right)^n\,,\quad f(z) = \frac{1}{1+\mu-\mu z}\,.}$$

The geometric distribution is a single-parameter distribution, the parameter being μ. All its moments must therefore be functions of μ alone.

1.5 Photon number distribution in blackbody radiation

A physical situation in which the geometric distribution occurs is provided by **blackbody radiation**, which is just a gas of photons in thermal equilibrium at some absolute temperature T. Here's a brief recapitulation of some background material.

Recall that a photon is a particle with zero rest mass, satisfying the energy-momentum relation $\varepsilon = c\,p$, where $p = |\mathbf{p}|$ is the magnitude of its momentum. The connection with electromagnetic waves is expressed via the relations

$$\varepsilon = h\nu,\ \mathbf{p} = \hbar\mathbf{k},\ \text{ and }\ \lambda = 2\pi/k.$$

Here ν is the frequency, λ is the wavelength, $\mathbf{k}$ is the wave vector, and $k = |\mathbf{k}|$ is the wave number. Hence the 'particle picture' relation $\varepsilon = c\,p$ is equivalent to the familiar 'wave picture' relation $c = \nu\,\lambda$.

A photon is specified not only by its momentum $\mathbf{p}$ (or wave number $\mathbf{k}$), but also its **spin S**. The spin quantum number of a photon is $S = 1$. Based on quantum mechanics, you might then expect the spin angular momentum along specified direction to have only $(2S + 1) = 3$ possible values, namely, $-\hbar$, 0 and $+\hbar$. However, the photon has zero rest mass, as already mentioned. Such a particle has only one possible speed in a vacuum, namely, the fundamental (or limiting) speed c.

- It is a consequence of relativistic quantum mechanics that a particle whose speed is always c (i.e., a zero-rest mass particle) has only *two* possible spin states, no matter what the quantum number S is.

The spin angular momentum of a photon is always directed either parallel or antiparallel to its momentum $\mathbf{p}$. That is, the quantity $(\mathbf{S}\cdot\mathbf{p})/(Sp)$ (called the **helicity** of the particle) can only take on the values $+1$ and -1, in units of $\hbar$. These correspond, in the wave language, to right circular polarization and left circular polarization, respectively. They are the only possible states of **polarization** of a single free photon.

Blackbody radiation is also called **thermal radiation**. It comprises photons of all wave vectors (and hence, all frequencies). It is also unpolarized, which means that it is an *incoherent superposition* of photons of both states of polarization for every wave vector. Since the spin quantum number of a photon is an integer ($S = 1$), photons obey **Bose-Einstein statistics**. Now consider all photons of a given wave vector $\mathbf{k}$ (and hence a given wave number k and frequency ν) and a given state of polarization (either left or right circular polarization) in blackbody radiation at a temperature T. We want to find the probability P_n that there are n such photons at any instant of time. We should really write $n(\mathbf{k}, +1)$ or $n(\mathbf{k}, -1)$ to indicate the fact that these are photons of a given wave vector and polarization state, but let's keep the notation simple.

There's a subtlety involved here that you must appreciate. A photon gas differs in two important ways from a gas of particles of nonzero rest mass in a container:

(i) In the case of a gas of particles with nonzero rest mass, collisions between the particles help maintain the thermal equilibrium of the gas. On the other hand, the mutual interaction between photons themselves is quite negligible. (There *is* a small photon-photon scattering cross-section induced by a quatum-electrodynamic process, but the effect of this interaction is extremely small at normal intensities of the radiation field.)

(ii) The number of massive particles in a closed container is constant. In stark contrast, the number of photons in a blackbody cavity fluctuates, because of their absorption and emission by the atoms in the walls of the radiation cavity. In fact, it is precisely this atom-photon interaction that is responsible for maintaining the thermal equilibrium of the radiation.

Although the photon number n fluctuates, all ensemble averages (in this case, the mean photon number and all its higher moments, etc.) are guaranteed to remain constant in time, because the system is in thermal equilibrium. In other words, the probability distribution of n must be independent of time. According to the rules of Bose-Einstein statistics, the mean number of photons of wave vector $\mathbf{k}$ and helicity equal to either $+1$ or -1 is given by

$$\langle n \rangle = \frac{1}{e^{\beta h\nu} - 1}\,,$$

where $\nu = ck/(2\pi)$, $\beta = (k_B T)^{-1}$, and k_B is Boltzmann's constant. Hence

$$e^{-\beta h\nu} = \frac{\langle n \rangle}{\langle n \rangle + 1}\,.$$

But we know that when a system is in thermal equilibrium at a temperature T, the probability that it has an energy E is proportional to $e^{-\beta E}$. The energy of n photons of frequency ν is just $nh\nu$. Hence the probability P_n is proportional to $e^{-n\beta h\nu}$. The constant of proportionality is easily determined by using the normalization condition $\sum_{n=0}^{\infty} P_n = 1$. We find

$$P_n = \left(1 - e^{-\beta h\nu}\right)\, e^{-n\beta h\nu}.$$

Substituting for $e^{-\beta h\nu}$ in terms of $\langle n \rangle$, we get the normalized geometric distribution

$$P_n = \frac{1}{\langle n \rangle + 1}\Big(\frac{\langle n \rangle}{\langle n \rangle + 1}\Big)^n \quad \text{where} \quad n = 0,\, 1,\, \ldots\,.$$

The probability distribution of the number of photons of a given wave vector and polarization in blackbody radiation is therefore a geometric distribution. Because of this connection with Bose-Einstein statistics, the geometric distribution itself is sometimes called the Bose-Einstein distribution, especially in the quantum optics literature concerned with photon-counting statistics.

★ **7.** Show that the variance and relative fluctuation of n are given, respectively, by

$$\mathrm{Var}\,(n) = \langle n\rangle\,\big(\langle n\rangle + 1\big) \quad \text{and} \quad \frac{\Delta n}{\langle n\rangle} = \Big(1 + \frac{1}{\langle n\rangle}\Big)^{1/2},$$

and that the k^{th} factorial moment of n is given by

$$\langle n(n-1)\,\ldots\,(n-k+1)\rangle = \langle n\rangle^k\,k!$$

The geometric distribution of the photon number in blackbody radiation is a direct consequence of (i) Bose-Einstein statistics, (ii) the absence of an interaction between photons, and (ii) the fact that the chemical potential of a photon gas is zero. Observe that the relative fluctuation in n is always greater than unity: in thermal radiation, the photon number has a large scatter about its mean value. As you will see shortly, this feature is in stark contrast to what happens in the case of coherent radiation. The same remark applies to the factorial moments as well.

1.6 From the binomial to the Poisson distribution

The **Poisson distribution** is a discrete probability distribution that occurs in a very large number of physical situations. It can be understood as a limiting case of the binomial distribution when the number of Bernoulli trials tends to infinity, while the probability of success in a single trial tends to zero, such that their product tends to a finite positive limit. That is,

$$\lim_{\substack{N\to\infty\\ p\to 0}} (N\,p) = \mu.$$

★ **8.** Show that, in this limit, the binomial distribution

$$P_n = \frac{N!}{(N-n)!\,n!}\;p^n\,(1-p)^{N-n} \quad (0 \le n \le N)$$

goes over into the Poisson distribution

$$P_n = e^{-\mu}\;\frac{\mu^n}{n!}\,, \quad \text{where} \;\; n = 0,\,1,\,\ldots\,.$$

(Eliminate p by setting it equal to μ/N. Use Stirling's formula for the factorial of a large number.)

The sample space of a Poisson-distributed random variable n is the infinite set of non-negative integers 0, 1, ... The parameter μ is nothing but the mean value of n, i.e., $\langle n \rangle = \mu$. The generating function of the Poisson distribution is

$$f(z) = \sum_{n=0}^{\infty} P_n \, z^n = e^{\mu(z-1)} \, .$$

The Poisson distribution is again a single-parameter distribution—there is only one parameter involved, namely, μ, which is the mean value itself. Hence all the higher moments of the random variable n must be expressible in terms of the mean value. The k^{th} factorial moment in this case is just

$$\langle n(n-1) \, \ldots \, (n-k+1) \rangle = \left. \frac{d^k f(z)}{dz^k} \right|_{z=1} = \mu^k \, .$$

In particular,

$$\text{Var}\,(n) = \langle n \rangle \;\; \text{for a Poisson distribution.}$$

The equality of the variance and the mean is a characteristic signature of the Poisson distribution. This relationship is a particular case of an even more special property of the Poisson distribution: all its cumulants are equal to μ. (We'll define cumulants shortly.)

Returning to the example of density fluctuations in an ideal gas, the passage to the limiting Poisson distribution has a physical interpretation. We have in this case $p = \rho v/N$, and the limit in which Np tends to a finite limit (which is clearly ρv) is just the thermodynamic limit: the number of molecules $N \to \infty$ and the volume $V \to \infty$ such that the ratio N/V tends to a finite value, the mean number density ρ. In this limit, the relative fluctuations in thermodynamic quantities vanish, and thermodynamics (which involves only average values of macroscopic observables) becomes exact. In this example, the probability that there are n gas molecules in a volume v becomes a Poisson distribution,

$$P_n = e^{-\rho v} \, \frac{(\rho v)^n}{n!} \, ,$$

with a mean value given by $\langle n \rangle = \rho v$, as you would expect.

1.7 Photon number distribution in coherent radiation

Another physical situation in which the Poisson distribution occurs is again provided by photons. In an ideal single-mode laser, the radiation is made up of photons of the same wave vector and polarization, and is said to be in a coherent state of the electromagnetic field. The properties of the quantum mechanical

linear harmonic oscillator enable us to describe coherent radiation quantum mechanically. What happens, in a nutshell, is as follows. When the electromagnetic field is *quantized*, each 'mode' of the field—that is, each component of a given frequency and state of polarization—behaves like a quantum mechanical linear harmonic oscillator. The **annihilation** and **creation operators** of the quanta (or photons) of each given wave vector and state of polarization are exactly like the lowering and raising operators (or ladder operators) a and $a^\dagger$ of the oscillator. They satisfy the canonical commutation relation $[a\,,\,a^\dagger] = I$. The operator $N = a^\dagger\,a$ is now the photon number operator. (Obviously, we should label a and $a^\dagger$ with the wave vector $\mathbf{k}$ and the state of polarization of the mode concerned. But we're concerned with just a single mode, and so let's keep the notation simple.) Its eigenvalues are, of course, the integers $n = 0,\, 1,\, \ldots$, with corresponding eigenvectors $|\,0\,\rangle\,,\,|\,1\,\rangle\,,\,\ldots$. The energy eigenstate $|\,n\,\rangle$ of the oscillator is equivalent to the **Fock state** of the electromagnetic field with exactly n photons of a given wave vector and polarization.

Every complex number α is a possible eigenvalue of the lowering operator a. The corresponding normalized eigenstate is a so-called **coherent state**, conveniently denoted by $|\alpha\rangle$. Thus, we have the eigenvalue equation

$$a\,|\alpha\rangle = \alpha\,|\alpha\rangle \quad \text{for any} \;\; \alpha \in \mathbb{C}.$$

★ **9.** Expand $|\alpha\rangle$ in the Fock basis as $|\alpha\rangle = \sum_{n=0}^{\infty} c_n\,|\,n\,\rangle$, and determine the coefficients $\{c_n\}$ by imposing the eigenvalue equation for a. You will need the basic relations $a\,|\,n\,\rangle = \sqrt{n}\,|n-1\rangle$. Choosing the phase of c_0 to be zero, show that the normalized coherent state is given by

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|\,n\,\rangle.$$

It is now very easy to determine the probability distribution P_n of the number of photons in the coherent state $|\alpha\rangle$. According to a basic rule of quantum mechanics, the **probability amplitude** that there are exactly n photons in the coherent state $|\alpha\rangle$ is given by the inner product $\langle n|\alpha\rangle$. The actual probability (that the number of photons in this state is n) is of course the square of the modulus of the probability amplitude. Using the orthonormality property of the set of Fock states, we get

$$\langle n|\alpha\rangle = \frac{e^{-\frac{1}{2}|\alpha|^2}\,\alpha^n}{\sqrt{n!}}\,,$$

Therefore

$$P_n \equiv |\langle n|\alpha\rangle|^2 = \frac{e^{-|\alpha|^2}\,|\alpha|^{2n}}{n!}\,.$$

This is a Poisson distribution with mean value $|\alpha|^2$. Hence the photon number distribution in ideal, single-mode coherent radiation is a Poisson distribution. The variance in the photon number is therefore equal to the mean value. The relative fluctuation in the photon number is $1/|\alpha|$. It decreases as the mean number of photons in the coherent state increases, in marked contrast with the case of thermal radiation (where it never falls below unity).

1.8 The negative binomial distribution

We've seen that the number distribution of photons of a given wave vector and state of polarization in coherent radiation is a Poisson distribution, while that in thermal radiation is a geometric distribution. The natural question to ask is: does there exist a *family* of distributions, of which the two distributions above are members? The **negative binomial distribution** provides such a family.

This question is of practical interest for the following reason. In general, radiation from a laser may contain more than one mode. But even in the case of a single mode, the coherent radiation is likely to be mixed with some thermal radiation as well. Photon-counting statistics enables us to analyze this admixture by quantifying the deviation from Poisson statistics. It's useful, for this purpose, to model the photon statistics of the admixture in terms of the negative binomial distribution.

The negative binomial distribution is a discrete distribution in which the random variable n takes the integer values 0, 1, ... *ad inf.*, with probability

$$P_n = \binom{N+n-1}{n} p^N q^n,$$

where $0 < p < 1$, $q = 1 - p$, and the parameter N is a positive integer. This distribution is therefore characterized by *two* parameters, p and N. It is immediately obvious that the case $N = 1$ reduces to the geometric distribution. The generating function of the distribution is

$$f_N(z) = \sum_{n=0}^{\infty} P_n z^n = p^N \sum_{n=0}^{\infty} \frac{(N+n-1)!}{(N-1)!\,n!} (qz)^n.$$

The series above can be summed by elementary methods, once it is recognized as just a binomial series. The result is

$$f_N(z) = p^N (1 - qz)^{-N}.$$

The form of $f(z)$ tells you why the distribution is called the negative binomial distribution.

★ **10.** Verify that the generating function of the negative binomial distribution is $f_N(z) = p^N(1-qz)^{-N}$. Show that the mean, variance, and relative fluctuation of n are given, respectively, by

$$\langle n\rangle = \frac{Nq}{p}\,, \quad \mathrm{Var}\,(n) = \frac{Nq}{p^2}\,, \quad \frac{\Delta n}{\langle n\rangle} = \frac{1}{\sqrt{(Nq)}}\,.$$

There exists a more general form of the negative binomial distribution. Here the parameter N is allowed to be any positive number, not necessarily an integer. The binomial coefficient in the definition of the distribution must then be expressed in terms of gamma functions, i.e.,

$$P_n = \binom{N+n-1}{n} p^N\, q^n = \frac{\Gamma(N+n)}{\Gamma(N)\, n!}\, p^N\, q^n.$$

This form enables us to see how the Poisson distribution arises as a limiting case of the negative binomial distribution, as follows.

★ **11.** Denote the mean value $\langle n\rangle$ by μ, so that $\mu = N(1-p)/p$. Hence $p = N/(N+\mu)$ and $q = \mu/(N+\mu)$. Substitute these expressions in the expression above for P_n, and pass to the limit $N \to \infty$. Use Stirling's approximation for the gamma functions, to arrive at the result

$$\lim_{N\to\infty}\; P_n = e^{-\mu}\; \frac{\mu^n}{n!}\,,$$

the Poisson distribution with mean value μ.

- The negative binomial distribution reduces to the geometric distribution for $N = 1$, and to the Poisson distribution as $N \to \infty$ keeping $\langle n\rangle\,(=\mu)$ finite.

As I have mentioned already, the negative binomial distribution is used in photon-counting statistics to characterize admixtures of thermal and coherent radiation. The family of negative binomial distributions *interpolates* between the geometric and Poisson distributions.

1.9 The sum of Poisson-distributed random variables

Often, one has to deal with a random variable that is the sum of two or more independent random variables, each of which is Poisson-distributed. What is the probability distribution of this sum? Let's first consider the case of two random variables. The more general case will turn out to be a simple extension of this

case, owing to a fundamental property of the Poisson distribution.

Let m and n be two independent random variables, each of which has a Poisson distribution, given by

$$P_1(m) = e^{-\mu}\ \frac{\mu^m}{m!} \quad \text{and} \quad P_2(n) = e^{-\nu}\ \frac{\nu^n}{n!}\,,$$

respectively. (I've used two different symbols, P_1 and P_2, for the two distributions in order to distinguish between them—one of them involves the parameter μ, while the other involves the parameter ν.) Let $s = m + n$ denote the sum of the two random variables. It is obvious that the sample space of s is again the set of non-negative integers. We want to find the probability distribution $P_{\rm sum}(s)$ of the sum s. This is obtained as follows: (a) multiply together the two individual probability distributions to get the joint probability (since the variables are independent of each other); (b) sum over all the possibilities for the individual random variables, subject to the constraint that the sum of m and n be equal to s. Thus,

$$P_{\rm sum}(s) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P_1(m)\, P_2(n)\, \delta_{m+n\,,\,s}\,.$$

★ **12.** Carry out the summations to obtain the result

$$P_{\rm sum}(s) = e^{-(\mu+\nu)}\ \frac{(\mu+\nu)^s}{s!}\,,\ \ s = 0,\, 1,\, \ldots\,.$$

Hence the sum of the two Poisson-distributed random variables m and n is also Poisson-distributed, with a mean value

$$\langle s \rangle = \mu + \nu = \langle m \rangle + \langle n \rangle.$$

It follows that

$$\mathrm{Var}\,(s) = \mathrm{Var}\,(m) + \mathrm{Var}\,(n) = \mu + \nu\,.$$

Further, the sum of any number of Poisson-distributed, independent random variables is also Poisson-distributed, with mean values and variances simply adding up.

1.10 The difference of two Poisson-distributed random variables

Consider again the pair of Poisson-distributed random variables m and n, with respective mean values μ and ν. Let $r = m - n$ be the *difference* of the two

random variables. We seek the probability distribution of r.

It is immediately obvious that r cannot have a Poisson distribution, because it can take on all possible integer values, including *negative* ones—its sample space is $\mathbb{Z}$. The probability distribution of r is given by

$$P_{\text{diff}}(r) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P_1(m)\, P_2(n)\, \delta_{m-n\,,\,r}\,.$$

Once again, the sum over n can be eliminated using the Kronecker delta to replace n by $m - r$. You must be careful about the limits of the subsequent summation over m. Draw the $(m\,,n)$ plane, mark the lattice points $m \geq 0\,,\ n \geq 0$, and note the set of points satisfying the constraint $m - n = r$. It is immediately obvious that two cases arise, corresponding to $r \geq 0$ and $r < 0$, respectively. Consider the case $r \geq 0$ first. The summation over m must obviously run from r upward, because n does not take on negative values. Hence the expression for $P_{\text{diff}}(r)$ reduces to

$$P_{\text{diff}}(r) = \sum_{m=r}^{\infty} P_1(m)\, P_2(m - r) \quad (r \geq 0).$$

Now consider $r < 0$. In this case all values of m from 0 upward contribute to the sum. Hence

$$P_{\text{diff}}(r) = \sum_{m=0}^{\infty} P_1(m)\, P_2(m + |r|) \quad (r < 0).$$

It remains to substitute the given Poisson distributions for P_1 and P_2 in the sums above, and to carry out the summations.

It turns out that the answer can be expressed in terms of the **modified Bessel function** of the first kind and of order l, denoted by $I_l(z)$. Recall that his function is defined, when l is not a negative integer, by the power series

$$I_l(z) = \sum_{m=0}^{\infty} \frac{1}{m!\,\Gamma(m + l + 1)}\, \left(\tfrac{1}{2}z\right)^{l+2m}\,.$$

The series converges for all finite values of $|z|$, so that $I_l(z)$ is an entire function of z. When $l = 0,\ 1,\ 2,\ \ldots$, the gamma function reduces to a factorial, and we have

$$I_l(z) = \sum_{m=0}^{\infty} \frac{1}{m!\,(m + l)!}\, \left(\tfrac{1}{2}z\right)^{l+2m}\,.$$

For negative integer values of l, the modified Bessel function is *defined* by the symmetry property

$$I_{-l}(z) \stackrel{\text{def.}}{=} I_l(z), \quad l \in \mathbb{Z}.$$

★ **13.** Use these relations to show that

$$P_{\text{diff}}(r) = e^{-(\mu+\nu)}\,(\mu/\nu)^{r/2}\,I_r(2\sqrt{\mu\,\nu}) \;\text{ for all } r \in \mathbb{Z}.$$

This distribution is also known as the **Skellam distribution**. Note that (i) it is a distribution on the *full* set of integers, and (ii) it is characterized by *two* parameters, μ and ν.

Recall that the generating function for the modified Bessel function of the first kind is given by

$$\sum_{r=-\infty}^{\infty} I_r(t)\,z^r = \exp\left\{\tfrac{1}{2}t\left(z+z^{-1}\right)\right\}.$$

You will recognise that this is a Laurent series that converges absolutely for all z in the annular region $0 < |z| < \infty$. It follows that the generating function of the Skellam distribution is given by

$$f(z) = \sum_{r=-\infty}^{\infty} P_{\text{diff}}(r) z^r = e^{-(\mu+\nu)}\,e^{\mu z+\nu z^{-1}}.$$

Thus $f(1) = 1$, verifying that $P_{\text{diff}}(r)$ is a correctly normalised distribution. It is obvious that the mean value of $r = m - n$ is given by $\langle r\rangle = \mu - \nu$.

★ **14.** Show that $\text{Var}\,(r) = \text{Var}\,(m-n) = \mu + \nu$. More generally, show that, if a and b are arbitrary real constants,

$$\text{Var}\,(am + bn) = a^2\,\mu + b^2\,\nu.$$

1.11 Cumulants and the cumulant generating function

The idea behind the variance of a random variable is generalized to the higher moments of the variable by means of the **cumulants** κ_r of a probability distribution. These quantities arise naturally as follows. The moment generating function of a random variable n $M(u)$ is the average value of the exponential e^{un}. The idea is to write this average value as the exponential of a function of u. The latter is, in general, a power series in u, whose coefficients are proportional to the cumulants of the random variable. We have

$$M(u) = \langle e^{un}\rangle = e^{K(u)}, \;\text{ where }\; K(u) \stackrel{\text{def.}}{=} \sum_{r=1}^{\infty} \frac{\kappa_r}{r!}\,u^r.$$

The quantities κ_r are called the cumulants of n, and $K(u)$ is the cumulant generating function. Note that $K(u)$ starts with the first-order term in u. Since $M(0) = 1$ for a normalized probability distribution, $K(0) = 0$. $K(u)$ is related to the moment generating function by

$$K(u) = \ln\, M(u).$$

The $r^{\rm th}$ cumulant is then found from the inversion formula

$$\kappa_r = \left[\frac{d^r K(u)}{du^r}\right]_{u=0}.$$

The cumulants can therefore be expressed in terms of the moments, and *vice versa*. The leading term in κ_r is $\langle n^r\rangle$, followed by terms involving the lower moments of n. The first cumulant is just the mean. The second and third cumulants are the corresponding central moments. Thus

$$\left.\begin{aligned}
&\kappa_1 = \mu = \langle n\rangle, \\
&\kappa_2 = \sigma^2 = \langle(\delta n)^2\rangle = \langle n^2\rangle - \langle n\rangle^2, \\
&\kappa_3 = \langle(\delta n)^3\rangle = \langle n^3\rangle - 3\langle n^2\rangle\,\langle n\rangle + 2\langle n\rangle^3\,.
\end{aligned}\right\}$$

The fourth cumulant is given by

$$\left.\begin{aligned}
\kappa_4 &= \langle(\delta n)^4\rangle - 3\langle(\delta n)^2\rangle^2 \\
&= \langle n^4\rangle - 4\langle n^3\rangle\,\langle n\rangle - 3\langle n^2\rangle^2 + 12\langle n^2\rangle\,\langle n\rangle^2 - 6\langle n\rangle^4\,.
\end{aligned}\right\}$$

A central problem in mathematical statistics is the *problem of moments*: given all the moments (and hence all the cumulants) of a distribution, can the distribution be reconstructed uniquely? I will not digress into this formal question here. In practice, however, the first four cumulants of a distribution provide a reasonable approximate description of the salient properties of the probability distribution of the random variable concerned.

The cumulant κ_r is a *homogeneous* function of order r in the following sense: if the random variable n is multiplied by a constant c, the $r^{\rm th}$ cumulant of cn is c^r times the corresponding cumulant of n. Observe that each κ_r starts with the $r^{\rm th}$ moment $\langle n^r\rangle$, but involves subtracting away contributions from the products of lower-order moments. This should remind you of the way in which we obtain spherical tensors from Cartesian tensors, to get irreducible sets of quantities under rotations. Something similar is involved here, too. For $r \geq 2$, the cumulants κ_r are invariant under *translations* of the random variable, i.e., under a change of variables $n \mapsto n + a$ where a is a constant. More than the moments themselves, or even the central moments, the different cumulants help characterize *distinct*

properties of a distribution. In general terms, the second cumulant (i.e., the variance) is a measure of the spread or dispersion of the random variable about its mean value. The third cumulant κ_3 measures the asymmetry or **skewness** of the distribution about the mean value. The fourth cumulant κ_4 characterizes, for a symmetric distribution, the extent to which the distribution deviates from the normal or Gaussian distribution.

★ **15.** Consider the discrete probability distributions discussed in the foregoing. Eliminate parameters wherever applicable in favor of μ, which represents the mean value (except in the case of the Skellam distribution). Show that the cumulant generating functions are given by

$$K(u) = \begin{cases} (\mu/p)\,\ln\,(1-p+p\,e^u) & \text{(binomial)} \\ -\,\ln\,(1+\mu-\mu e^u) & \text{(geometric)} \\ \mu\,(e^u-1) & \text{(Poisson)} \\ \mu\,(e^u-1)-\nu(1-e^{-u}) & \text{(Skellam)} \\ N\,\ln\,N-N\,\ln\,(N+\mu-\mu\,e^u) & \text{(negative binomial).} \end{cases}$$

Find the first four cumulants explicitly in each of the cases above.

2 Continuous probability distributions

2.1 Probability density and cumulative distribution

We move on now to (real) random variables whose sample space is a continuum. The probability that a continuous random variable takes any particular, precisely specified, value in a continuous set is actually zero, in general. (This is true unless there is a δ-function contribution to the probability density at that point, as we'll see below.) The reason is that a single point, specified to infinite precision, is a set of measure zero in a continuum. For *continuous* random variables, therefore, we must speak of a **probability density function** (PDF). It's useful to distinguish between a continuous random variable *per se* and the *values* it can take, by using different (but related) symbols for the two—e.g., upper case and lower case letters such as X and x, respectively. Wherever it is preferable to do so, I will make this distinction. Suppose the sample space of a random variable X is $(-\infty, \infty)$, and $p(x)$ is its PDF. Then:

(i) $p(x)\,dx$ is the probability that the random variable X has a value in an *infinitesimal range* dx at the value x, i.e., in the range $(x,\, x + dx)$.

(ii) $p(x) \geq 0 \quad$ for all $\quad x$.

(iii) $\int_{-\infty}^{\infty} dx\, p(x) = 1$.

Note that $p(x)$ itself does not have to be less than unity. In fact, it can even become unbounded in the domain of x. But it must be integrable, because of condition (iii) above.

The **cumulative distribution function** (CDF), also called simply the distribution function, is the total probability that the random variable X is less than, or equal to, any given value x. Thus, in the case of a sample space $(-\infty\,,\, \infty)$,

$$\Pr\,(X \leq x) \equiv F(x) = \int_{-\infty}^{x} dx'\, p(x').$$

$F(x)$ is a *non-decreasing*, non-negative function of x that satisfies $F(-\infty) = 0$, $F(\infty) = 1$. It is obvious that $dF/dx = p(x)$.

The CDF is a more general concept than the PDF. All probability distributions need not necessarily have well-behaved functions as corresponding probability densities. For instance, there may be some specific value of the random variable, say x_0, with a nonzero probability α, say, of its occurrence. In the heuristic approach we have taken, we would simply say that the PDF $p(x)$ has a 'spike' of the form $\alpha\,\delta(x - x_0)$. In a more rigorous treatment, care must be exercised in handling such singularities. A δ-function spike in $p(x)$ leads to a *finite*

jump or step in $F(x)$. This is one reason why it is found to be more convenient, mathematically, to deal with the CDF rather than the PDF.

As I've mentioned, the CDF is frequently called the distribution function. In physical applications, one often goes a step further and refers to the PDF itself as the 'probability distribution' or just the 'distribution'. While this terminology is loose, it is very common in the physics literature. No confusion should arise, as matters will be clear clear from the context. Note that

(i) $p(x)$ has the physical dimensions of $[x]^{-1}$, while $F(x)$ is dimensionless.

(ii) $F(x)$ can never exceed unity (but there is no such restriction on $p(x)$, of course).

2.2 The moment and cumulant generating functions

Probability distributions themselves are not measurable or 'physical' quantities. All that we can measure directly, or find, are various averages. For a single random variable, this is the set of its moments $\langle X^r \rangle$, $r = 1, 2, \ldots$. As in the case of a discrete-valued random variable, we define the moment generating function

$$M(u) = \langle e^{uX} \rangle = 1 + \sum_{r=1}^{\infty} \frac{\langle X^r \rangle}{r!}\, u^r, \quad \text{so that} \quad \langle X^r \rangle = \left[\frac{d^r M(u)}{du^r} \right]_{u=0} .$$

In particular, the mean value $\langle X \rangle \equiv \mu$ is the coefficient of u in the power series expansion of $M(u)$.

More useful than the moments are the **central moments** of the random variable. These are the moments of the *deviation* of X from its mean value, namely,

$$\langle (\delta X)^r \rangle = \langle (X - \mu)^r \rangle.$$

The first of these is zero, by definition. The second is the variance,

$$\sigma^2 \equiv \mathrm{Var}\,(X) = \langle (X - \mu)^2 \rangle = \langle X^2 \rangle - \mu^2.$$

As you know, the variance of a random variable is always positive (it vanishes if and only if the variable is deterministic). It provides the leading measure of the scatter of a random variable about its mean value.

The generating function $f(z)$ corresponding to the distribution of X is just

$$f(z) = \int dx\, p(x)\, z^x = \langle z^X \rangle.$$

It is obvious that f and M are related to each other according to

$$M(u) = f(e^u).$$

As in the case of a discrete random variable, the cumulant generating function $K(u)$ is related to $M(u)$ according to

$$M(u) = e^{K(u)}, \quad \text{or} \quad K(u) = \ln\, M(u).$$

The r^{th} cumulant is given by $\kappa_r = \left[d^r K(u)/du^r\right]_{u=0}$.

★ **16.** I've already stated that the cumulants κ_r for $r \geq 2$ remain unchanged under a shift of the random variable by a constant. Here's an example to illustrate this property. Let n be a Poisson-distributed random variable with mean value μ. Consider the random variable $X = an + b$, where a and b are arbitrary positive constants (not necessarily integers).

(a) Find the probability distribution (or the PDF) of X, and its moment generating function.

(b) Hence find the cumulant generating function of X, and show that its cumulants are given by $\kappa_1 = a\mu + b$ and $\kappa_r = a^r\,\mu$ for $r \geq 2$. This shows that κ_r is unchanged if the random variable n is merely shifted by a constant to $n+b$.

2.3 The characteristic function

A very useful quantity associated with the PDF $p(x)$ of a random variable X is its Fourier transform $\widetilde{p}(k)$, given by

$$\widetilde{p}(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx}\, p(x).$$

$\widetilde{p}(k)$ is called the **characteristic function** of the distribution. It is closely related to the moment generating function of the random variable: $\widetilde{p}(k)$ can also be regarded as the expectation value of $\exp(-ikX)$, so that

$$\widetilde{p}(k) = \langle e^{-ikX} \rangle = M(-ik) = f(e^{-ik}).$$

(Recall that M is the moment generating function and f is the generating function associated with the random variable.) Thus $\langle X^r \rangle / r!$ is the coefficient of $(-ik)^r$ in the power series expansion of $\widetilde{p}(k)$. The characteristic function carries the same information about the random variable as does its PDF. The significance and utility of the characteristic function will become clear shortly.

A function of $k \in (-\infty, \infty)$ has to have some special features in order to qualify as the characteristic function $\widetilde{p}(k)$ of a probability distribution:

(i) $\widetilde{p}(0)$ must be equal to unity, to ensure that the random variable X is a proper random variable, with a normalized probability distribution.

(ii) The inverse Fourier transform of $\widetilde{p}(k)$ must be a real, non-negative function of x, in order to be an acceptable PDF $p(x)$.

The second of these requirements, in particular, places considerable restrictions on the function $\widetilde{p}(k)$.

2.4 The additivity of cumulants

A crucial property of cumulants emerges when we consider *sums* of independent random variables. Let X_1 and X_2 be independent random variables, each with a sample space comprising all real numbers, and with PDFs $p_1(x_1)$ and $p_2(x_2)$, respectively. Let $X = X_1 + X_2$ be their sum. The PDF of X is given by

$$p(x) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, p_1(x_1)\, p_2(x_2)\, \delta\big(x - (x_1 + x_2)\big) = \int_{-\infty}^{\infty} dx_1\, p_1(x_1)\, p_2(x - x_1).$$

But this is just a convolution of the PDFs of x_1 and x_2. Hence, by the convolution theorem for Fourier transforms, the Fourier transform of the function p is just the product of the Fourier transforms of the functions p_1 and p_2. In other words, the corresponding characteristic functions are related according to

$$\widetilde{p}(k) = \widetilde{p}_1(k)\, \widetilde{p}_2(k), \quad \text{or} \quad M(-ik) = M_1(-ik)\, M_2(-ik),$$

where M_1 and M_2 are the moment generating functions of X_1 and X_2. Taking logarithms, it follows at once that the corresponding cumulant generating functions simply add up:

$$K(-ik) = K_1(-ik) + K_2(-ik).$$

As a consequence, the r^{th} cumulant of $X_1 + X_2$ is the sum of the r^{th} cumulants of X_1 and X_2, for every positive integer r. It is obvious that this result can be extended immediately to the sum (more generally, to any linear combination) of any number of independent random variables. Unlike the cumulants, none of the moments higher than the first moment (or mean value) has this property of additivity. This is a major advantage that the cumulants of random variables enjoy over the corresponding moments.

2.5 The Gaussian (or normal) distribution

The **Gaussian distribution**, also called the normal distribution, plays a central role in statistics. Its appearance in applications is also ubiquitous. The distribution of a continuous random variable X that takes values in $\in (-\infty\,,\,\infty)$ is

Gaussian if its normalized PDF is given by

$$p(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\Big\{ - \frac{(x-\mu)^2}{2\sigma^2} \Big\},$$

where μ is any real number and σ is a positive number. The parameter μ is the mean value of X, σ^2 is its variance, and σ is its standard deviation:

$$\langle X \rangle = \mu, \ \langle (X-\mu)^2 \rangle = \sigma^2.$$

In statistics, the notation $\mathcal{N}(\mu, \sigma^2)$ is often used to denote the normal distribution. $\mathcal{N}(0,1)$ is the *standard* normal distribution. The Gaussian PDF is an example of a *symmetric, unimodal* density: it has a single peak or *mode*, which also happens to be located at $x = \mu$, and the plot of $p(x)$ is symmetric about this point. The variance σ^2 is a direct measure of the width of the Gaussian. The full-width-at-half-maximum (FWHM) is given by

$$\mathrm{FWHM}\big|_{\mathrm{Gaussian}} = 2\,(2\,\ln\,2)^{1/2}\sigma \simeq 2.355\,\sigma.$$

Here is a physical instance in which the Gaussian distribution appears. Recall the number fluctuations in a classical ideal gas. We've seen how, in the thermodynamic limit, the binomial distribution goes over into the Poisson distribution: the probability that a volume v of the gas has n molecules is given by $P_n = e^{-\langle n \rangle}\, \langle n \rangle^n / n!$, where $\langle n \rangle = \rho v$ and ρ is the mean number density of the gas. At normal temperature and pressure, the mean number of molecules in a cubic metre of air is of the order of 10^{22}. This is so large compared to unity that we may regard the *deviation* $X = n - \langle n \rangle$ in the number of molecules around its mean value to be an essentially continuous variable. The distribution of X is then a Gaussian with zero mean, with a variance that remains equal to $\langle n \rangle$.

★ 17. Establish this result, as follows. Consider the logarithm of P_n. Use Stirling's formula for $\ln\, n!$, replace n by $\langle n \rangle + x$, and P_n by the PDF $p(x)$. Use the fact that

$$\ln\,\big(\langle n \rangle + x\big) = \ln \langle n \rangle + \ln\,\Big(1 + \frac{x}{\langle n \rangle}\Big) \simeq \ln \langle n \rangle + \frac{x}{\langle n \rangle} - \frac{x^2}{2\langle n \rangle^2}\,.$$

After the cancellation of various terms, show that $\ln\, p(x)$ is proportional to $-x^2/(2\langle n \rangle)$, to leading order. Hence X has the normal distribution $\mathcal{N}(0, \langle n \rangle)$.

The cumulative distribution function of a Gaussian random variable X is given by

$$\Pr\,(X \le x) = F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} dx'\, \exp\Big\{ - \frac{(x'-\mu)^2}{2\sigma^2} \Big\}.$$

We have $F(-\infty) = 0$ and $F(\infty) = 1$, as required. Further, $F(\mu) = \frac{1}{2}$. Recall that the so-called **error function** is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty dt\, e^{-t^2}.$$

In terms of the error function, the CDF of a Gaussian random variable is given by

$$F(x) = \frac{1}{2} \left\{ 1 + \mathrm{erf}\, \left(\frac{x-\mu}{\sqrt{2\sigma^2}} \right) \right\}.$$

2.6 Moments and cumulants of a Gaussian distribution

Owing to the symmetry of the Gaussian PDF about the mean value, all the *odd* moments of the deviation from the mean value (that is, all the odd central moments) vanish identically:

$$\langle (X-\mu)^{2l+1} \rangle = \int_{-\infty}^{\infty} dx\, (x-\mu)^{2l+1}\, p(x) = 0, \quad l = 0,\, 1,\, \ldots\, .$$

On the other hand, all the *even* central moments of a Gaussian random variable are determined completely in terms of σ^2, i.e., in terms of the variance of the distribution.

★ **18.** Show that, for $l = 0,\, 1,\, \ldots$,

$$\langle (X-\mu)^{2l} \rangle = \int_{-\infty}^{\infty} dx\, (x-\mu)^{2l}\, p(x) = \frac{(2\sigma^2)^l}{\sqrt{\pi}}\, \Gamma(l+\tfrac{1}{2}) = \frac{(2l)\,!}{2^l\, l!}\, \sigma^{2l}.$$

Note, in particular, that

$$\langle (\delta X)^4 \rangle = 3\, \sigma^4 = 3\, \langle (\delta X)^2 \rangle^2.$$

The significance of this fact will become clear shortly.

The moment generating function for the Gaussian distribution is

$$M(u) = \langle e^{u\, X} \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} dx\ \exp \left\{ -\ \frac{(x-\mu)^2}{2\sigma^2} + u\, x \right\}$$

$$= \exp\, \left(\mu\, u + \tfrac{1}{2} \sigma^2\, u^2 \right).$$

The characteristic function of the Gaussian distribution is therefore

$$\widetilde{p}(k) = M(-ik) = \exp\, \left(-i\mu\, k - \tfrac{1}{2} \sigma^2\, k^2 \right).$$

The corresponding cumulant generating function is just

$$K(u) = \ln\, M(u) = \mu\, u + \tfrac{1}{2} \sigma^2\, u^2,$$

a quadratic function of u. A fundamental property follows at once:

- *All the cumulants of a Gaussian distribution higher than the second cumulant vanish identically.*

In particular, the fourth cumulant κ_4 is identically equal to zero for a Gaussian. This brings us to the significance of κ_4 for a general distribution that is symmetric about its mean value. The dimensionless ratio

$$\frac{\kappa_4}{\kappa_2^2} = \frac{\langle (\delta x)^4 \rangle - 3 \langle (\delta x)^2 \rangle^2}{\langle (\delta x)^2 \rangle^2}$$

is called the **excess of kurtosis**. This quantity is exactly zero for a Gaussian distribution. A positive excess of kurtosis (a **leptokurtic distribution**) implies that the fourth central moment dominates over the square of the second central moment. Hence larger values of $|x - \mu|$ contribute more significantly than they do in a Gaussian. In qualitative terms, this means that the PDF has a *leaner* peak than a Gaussian, and fatter 'tails' on either side of the peak. On the other hand, a negative excess of kurtosis (a **platykurtic distribution**) implies that the weight of large values of $|x - \mu|$ is less than what it would be in the case of a Gaussian. Hence the peak is *broader* than that of a Gaussian, while the tails on either side of the peak are thinner.

The cumulant generating function $K(u)$ is a quadratic in u for a Gaussian distribution. Hence κ_r vanishes identically for all $r \geq 3$ in this case. It is natural to ask: Are there continuous distributions for which $K(u)$ is a cubic, or quartic, or some polynomial of *finite* degree $l \geq 3$, so that all cumulants with $r \geq l + 1$ would vanish identically? Interestingly, the answer is that there are *no* such distributions.

2.7 Simple functions of a Gaussian random variable

An arbitrary *function* of a Gaussian random variable will, in general, have a PDF that looks quite different from a Gaussian. Here's a simple example. Let the random variable X have a distribution $\mathcal{N}(0, \sigma^2)$, so that its normalized PDF is

$$p(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\, e^{-x^2/(2\sigma^2)}.$$

What is the normalized PDF $\rho(\xi)$ of the random variable X^2? The sample space of X^2 is, of course, just the half-line $0 \leq \xi < \infty$. If the mapping $x \mapsto \xi$ had been one-to-one, then we could have used the formal identity $p(x)\,|dx| = \rho(\xi)\,|d\xi|$ to find $\rho(\xi)$. This identity is nothing but the following statement: if x lies in the range $(x,\, x + dx)$, then ξ lies in the range $(\xi,\, \xi + d\xi)$, and the corresponding probabilities are obviously equal to each other. But both $-x$ and $+x$ correspond to the same $\xi = x^2$. Hence there is a *two*-fold contribution to $\rho(\xi)$. This gives an

extra factor of 2, and we have

$$\rho(\xi) = 2\,p\big(\sqrt{\xi}\,\big)\,\left|\frac{dx}{d\xi}\right| = \frac{1}{\sqrt{(2\pi\sigma^2\,\xi)}}\,e^{-\xi/(2\sigma^2)}.$$

Alternatively, you could also use the formal expression

$$\rho(\xi) = \int_{-\infty}^{\infty} dx\,p(x)\,\delta(x^2-\xi) = \int_{-\infty}^{\infty} dx\,p(x)\,\frac{\delta(x+\sqrt{\xi})+\delta(x-\sqrt{\xi})}{2\sqrt{\xi}}\,,$$

to arrive at the same result. Note that $\rho(\xi)$ is properly normalized according to $\int_0^\infty d\xi\,\rho(\xi) = 1$.

A physical example: The most well-known example of a Gaussian distribution in physics is, of course, the Maxwellian distribution of velocities in a classical ideal gas in thermal equilibrium at an absolute temperature T. Each Cartesian component of the velocity of a molecule of mass m has a Gaussian PDF, with zero mean and a variance equal to k_BT/m. Thus

$$p^{\rm eq}(v_x) = \Big(\frac{m}{2\pi k_BT}\Big)^{1/2}\exp\Big(-\frac{mv_x^2}{2k_BT}\Big), \quad (-\infty < v_x < \infty)$$

and similarly for v_y and v_z. Correspondingly, the PDF $\rho^{\rm eq}(v)$ of the speed v of a molecule is given by **Maxwell-Boltzmann distribution**

$$\rho^{\rm eq}(v) = \Big(\frac{m}{2\pi k_BT}\Big)^{3/2}4\pi v^2\,\exp\Big(-\frac{mv^2}{2k_BT}\Big), \quad (0 \le v < \infty).$$

★ 19. Show that the PDF $\phi^{\rm eq}(\varepsilon)$ of the (kinetic) energy ε of a molecule is given by

$$\phi^{\rm eq}(\varepsilon) = \frac{2}{\sqrt{\pi}\,(k_BT)^{3/2}}\;\varepsilon^{1/2}\;e^{-\varepsilon/(k_BT)}, \quad (0 \le \varepsilon < \infty).$$

This PDF is an example of a **gamma distribution**. The latter is a two-parameter distribution of a random variable taking values in $[0,\infty)$, with a PDF proportional to $x^{a-1}\,e^{-x/b}$, where a (the *shape factor*) and b (the *scale parameter*) are positive constants. In the example at hand, $a = \frac{3}{2}$, $b = k_BT$.

2.8 Linear combinations of Gaussian random variables

The statistical properties of *sums* (more generally, of linear combinations) of independent Gaussian random variables are of considerable interest. Consider the simplest case first. Let $X_1\,,\,X_2\,,\ldots\,,X_n$ be n independent, identically-distributed

random variables (usually abbreviated as *iidrv*), each with a distribution $\mathcal{N}(\mu, \sigma^2)$. Then the random variable

$$Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}}$$

has the distribution $\mathcal{N}(0, 1)$, i.e., Z_n is is also a Gaussian random variable, with zero mean and unit variance.

The 'brute force' way to derive this result would be to compute the PDF of Z_n as follows: Take the product of the n individual Gaussian PDFs $p(x_i)$ of the the random variables X_i , multiply this by $\delta(z - z_n)$ where $z_n = \sum_{i=1}^{n}(x_i - \mu)/(\sigma\sqrt{n})$ (to take care of the definition of Z_n), and integrate the result over all the n variables x_i . It would be foolish to use the δ-function to carry out one of the integrations, because you would still be left with an $(n - 1)$-fold integral. A better way is to write the Fourier representation for the δ-function. This would immediately factorize the argument of the δ-function into a product, and each of the resulting Gaussian integrals can be evaluated. The result is the PDF $(2\pi)^{-1/2}\, e^{-z^2/2}$, which corresponds to a standard normal distribution.

But there's a simpler way to write down the answer. Recall the additivity property of the cumulants (or their generating function) for independent random variables, and the fact that they are invariant under a shift of the random variable by a constant! The cumulant generating function of each $(X_i - \mu)$ is $\frac{1}{2}\sigma^2 u^2$. It follows that the cumulant generating function of Z_n is given by $\frac{1}{2}n\sigma^2 u^2/(n\sigma^2) = \frac{1}{2}u^2$. Comparing this with the general expression for the cumulant generating function of a Gaussian distribution, we conclude that Z_n has a Gaussian distribution with zero mean and unit variance, $\mathcal{N}(0, 1)$.

A more general version of this result is as follows. Let the mean and variance of the Gaussian random variable X_i be μ_i and σ_i^2, respectively. Consider the linear combination

$$\xi_n = \sum_{i=1}^{n} a_i\, X_i\,,$$

where the constants a_i are real numbers. It follows from the additivity of the cumulants κ_1 and κ_2 that the mean and variance of ξ_n are given, respectively, by

$$\langle \xi_n \rangle = \sum_{i=1}^{n} a_i\, \mu_i \quad \text{and} \quad \langle (\xi_n - \langle \xi_n \rangle)^2 \rangle = \langle (\delta\xi_n)^2 \rangle = \sum_{i=1}^{n} a_i^2\, \sigma_i^2.$$

★ 20. Show that the random variable $\chi_n = (\xi_n - \langle \xi_n \rangle)/\langle (\delta\xi_n)^2 \rangle^{1/2}$ has the distribution $\mathcal{N}(0, 1)$.

2.9 The Central Limit Theorem

One of the most important theorems of statistics is the Central Limit Theorem. Widely regarded as the 'crown jewel' of the subject of probability and statistics, the theorem is actually a generic name for a class of convergence theorems in statistics. What is referred to here is the most common of these results:

Let $X_1\,, X_2\,, \ldots\,, X_n$ be *iidrv* with mean μ and variance σ^2. They need not be Gaussian random variables! Then, in the limit $n \to \infty$, the probability distribution of the random variable $Z_n = (X_1 + \ldots + X_n - n\mu)/(\sigma\sqrt{n})$ tends to a Gaussian distribution with zero mean and unit variance.

A Gaussian is therefore a **limit law** or **limit distribution** in this sense. Several of the conditions stated above can be relaxed without affecting the validity of the theorem. For instance, under certain conditions, the random variables need not be identically distributed. The crucial requirement, however, is the finiteness of the mean and variance of each of the random variables making up the sum. The Central Limit Theorem helps us understand why the Gaussian distribution occurs so frequently in all physical applications. In very broad, qualitative terms:

- When a large number of independent and uncorrelated effects contribute to a cause, one may expect the resultant, suitably shifted and re-scaled, to have a normal distribution.

This is essentially how the Maxwellian distribution of velocities arises in a gas in thermal equilibrium.

★ **21.** A random variable X is *uniformly distributed* in the unit interval $[0, 1]$. That is, its PDF is given by

$$p(x) = \begin{cases} 1 & \text{for} \quad 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Another, independent, random variable Y is also uniformly distributed in $[0, 1]$. Show that the PDF $\rho_2(z)$ of their sum $Z = X + Y$ is given by

$$\rho_2(z) = \begin{cases} z & \text{for} \quad 0 \leq z \leq 1 \\ 2 - z & \text{for} \quad 1 < z \leq 2. \end{cases}$$

2.10 An explicit illustration of the Central Limit Theorem

The example just considered shows how the uniform distribution in x and y leads to a tent-shaped PDF for their sum. As one keeps adding more such variables,

the PDF of the resultant acquires more and more bends, and approximates a Gaussian ever more closely. Here is how the Gaussian emerges as a limiting distribution in this instance.

Let X_1 , ... , X_n be *iidrv*, each distributed uniformly in $[0, 1]$. The mean value of each variable is obviously $\mu = \frac{1}{2}$, while its variance is $\sigma^2 = \langle x_i^2 \rangle - \mu^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$. Now consider the random variable

$$Z_n = \frac{X_1 + \ldots + X_n - \frac{1}{2}n}{\sqrt{n/12}}\,.$$

We're going to show that, as $n \to \infty$, the distribution of Z_n tends to the standard normal distribution $\mathcal{N}(0, 1)$. We'll do this in two slightly different (but equivalent) ways.

(i) Since the PDF of each X_i is just unity, the PDF of Z_n is

$$\rho_n(z) = \int_0^1 dx_1 \ldots \int_0^1 dx_n\, \delta(z - z_n),$$

where $z_n = \sum_1^n \left(x_i - \frac{1}{2}\right)/\sqrt{n/12}$. Using the Fourier representation

$$\delta(z - z_n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ik(z - z_n)},$$

we get

$$\rho_n(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ikz}\, e^{ik\sqrt{3n}}\, I^n,$$

where the integral

$$I = \int_0^1 dx\, e^{-ikx\sqrt{12/n}}.$$

Simplifying, we get

$$\rho_n(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ikz} \left\{ \frac{\sin\left(k\sqrt{3/n}\,\right)}{k\sqrt{3/n}} \right\}^n.$$

It is now easy to pass to the limit $n \to \infty$. Expand the sine in its power series, and note that

$$\lim_{n\to\infty} \left\{ \frac{\sin\left(k\sqrt{3/n}\right)}{k\sqrt{3/n}} \right\}^n = \lim_{n\to\infty} \left(1 - \frac{k^2}{2n}\right)^n = e^{-\frac{1}{2}k^2}.$$

Therefore

$$\lim_{n\to\infty}(z) \equiv \rho(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ikz}\, e^{-\frac{1}{2}k^2} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2},$$

a Gaussian PDF with zero mean and unit variance, as asserted.

(ii) Alternatively, we can arrive at the characteristic function of Z_n with a little less work, by the following argument. Since each X_i is uniformly distributed in $[0,1]$ with a PDF equal to unity, the random variable $Y_i = \sqrt{12}\,\big(X_i - \frac{1}{2}\big)/\sqrt{n}$ is uniformly distributed in $[-\sqrt{3/n}\,,\ \sqrt{3/n}]$, with a constant PDF equal to $\sqrt{n/12}$. Hence the moment generating function of any Y_i is given by

$$M_Y(u) = \sqrt{(n/12)} \int\limits_{-\sqrt{3/n}}^{\sqrt{3/n}} dy\, e^{uy} = \frac{\sinh\,(u\sqrt{3/n})}{u\sqrt{3/n}}.$$

The cumulant generating function of any Y_i is given by $\ln\, M_Y(u)$. Hence the cumulant generating function of $Z_n = \sum_1^n Y_i$, which is a sum of n independent random variables, is just $n\, \ln\, M_Y(u) = \ln\,[M_Y(u)]^n$. Exponentiating this quantity, the moment generating function of Z_n is

$$M_{Z_n}(u) = \Big\{\frac{\sinh\ u\sqrt{3/n}}{u\sqrt{3/n}}\Big\}^n.$$

The characteristic function of Z_n is therefore

$$\widetilde{\rho}_n(k) = M_{Z_n}(-ik) = \Big\{\frac{\sin\,(k\sqrt{3/n})}{k\sqrt{3/n}}\Big\}^n.$$

Taking the inverse Fourier transform gives us $\rho_n(z)$. The rest of the derivation proceeds as before.

★ **22.** Work out explicitly the steps of the derivations outlined above.

2.11 From random flights to diffusion

The manner in which a random walk goes over into the diffusion process in the limit of zero step size in both time and space serves as another beautiful illustration of the Central Limit Theorem.

In order to be specific, and to be closer to a physical situation, let's consider random flights of the following kind, in three-dimensional space of infinite extent. (For ease of identification, I'll use the term random *walk* for random motion on discrete spatial lattices, and random *flight* for random motion with finite-sized steps in a continuous space.) The walker starts at an arbitrary origin, and at the end of each time step τ, takes a step of fixed length l in an *arbitrary* direction in space. This assumption can be relaxed to include a distribution of step lengths,

but I'll not do so here. Each successive step is taken independently, and is uncorrelated with the steps preceding it. We ask for the normalized probability density function $p(\mathbf{r}, n\tau)$ of the position vector $\mathbf{r}$ of the walker at the end of n time steps, i.e., at time $n\tau$. I have written $p(\mathbf{r}, n\tau)$ rather than $p(\mathbf{r}, n)$ in order to make the time dependence explicit. The objective is to calculate this quantity, and to show that it reduces to the fundamental Gaussian solution of the diffusion equation when the limits $n \to \infty$, $l \to 0$ and $\tau \to 0$ are taken *in a specific manner*.

Let the successive steps of the walker be given by the (randomly directed) vectors $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n$, so that

$$\mathbf{r} = \sum_{j=1}^{n} \mathbf{R}_j .$$

As usual, we use the symbol $\langle \cdots \rangle$ to denote the statistical average. In the present instance, this is an average over all 'realizations' of the random flight, i.e., over *all possible configurations of the set of steps*. In effect, it is as if we have a chain with n straight, rigid segments, with loose-jointed hinges connecting every pair of adjacent segments. As each step is equally likely to be in any direction, it is obvious that the average value of the *vector* representing any step vanishes, i.e., $\langle \mathbf{R}_j \rangle = 0$. Hence the mean displacement vector $\langle \mathbf{r} \rangle = 0$, as you would expect intuitively. Further, using the fact that the magnitude of every step is equal to l, we have

$$\langle \mathbf{r} \cdot \mathbf{r} \rangle \equiv \langle r^2 \rangle = nl^2 + l^2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \langle \cos \theta_{ij} \rangle,$$

where θ_{ij} is the angle between the vectors $\mathbf{R}_i$ and $\mathbf{R}_j$. But $\langle \cos \theta_{ij} \rangle = 0$, because the angle between the two vectors is as likely to have a value θ as it is to have a value $(\pi - \theta)$, and $\cos(\pi - \theta) = -\cos \theta$. The contributions from the two possibilities add up to zero. It follows that

$$\langle r^2 \rangle = nl^2.$$

Since $\langle \mathbf{r} \rangle = 0$, $\langle r^2 \rangle$ is actually the variance of the *displacement* in an n-step random flight. (It is *not* the variance of the end-to-end *distance*, because the mean distance $\langle r \rangle$ is not zero.) As n is proportional to the duration t of the walk, you can already see the emergence of the *linear growth in time* of the variance of the displacement. This behavior is characteristic of **diffusive motion**. Note, in particular, that this result for the variance is actually *independent* of the number of dimensions of the space in which the random flight occurs!

Let's return to the calculation of the PDF $p(\mathbf{r}, n\tau)$. It turns out to be easier to compute its Fourier transform $\widetilde{p}(\mathbf{k}, n\tau)$ first. The latter is just the characteristic function of the random variable $\mathbf{r}$. Recall the multiplicative property of

characteristic functions for a sum of independent random variables. Since $\mathbf{r}$ is the sum of the independent random variables $\mathbf{R}_j$ $(1 \leq j \leq n)$, it follows that $\widetilde{p}\,(\mathbf{k}, n\tau)$ must be the product of the characteristic functions of the individual steps. It is instructive to see this explicitly, as follows. We have

$$\widetilde{p}(\mathbf{k}, n\tau) = \int d^3r \; e^{-i\mathbf{k}\cdot\mathbf{r}} \; p(\mathbf{r}, n\tau) \quad \Longleftrightarrow \quad p(\mathbf{r}, n\tau) = \frac{1}{(2\pi)^3} \int d^3k \; e^{i\mathbf{k}\cdot\mathbf{r}} \; \widetilde{p}(\mathbf{k}, n\tau).$$

But $\widetilde{p}(\mathbf{k}, n\tau)$ can be interpreted as the average value of the quantity $e^{-i\mathbf{k}\cdot\mathbf{r}}$ over all possible n-step random walks. Therefore we have

$$\widetilde{p}(\mathbf{k}, n\tau) = \left\langle e^{-i\mathbf{k}\cdot\mathbf{r}} \right\rangle = \left\langle e^{-i\sum_{j=1}^{n} \mathbf{k}\cdot\mathbf{R}_j} \right\rangle = \left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_1} \; \ldots \; e^{-i\mathbf{k}\cdot\mathbf{R}_n} \right\rangle.$$

A great simplification occurs now, *because the individual steps are completely independent of each other*. Hence the average value of the product in the final step above becomes equal to the product of average values, and we have

$$\widetilde{p}(\mathbf{k}, n\tau) = \left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_1} \right\rangle \left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_2} \right\rangle \cdots \left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_n} \right\rangle = \prod_{j=1}^{n} \left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_j} \right\rangle.$$

Now let $p(\mathbf{r}, \tau) \equiv p_1(\mathbf{r})$ denote the PDF of a *single* step. Since this PDF is the *same* for each of the steps, we have

$$\left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_j} \right\rangle = \int d^3\mathbf{R}_j \, e^{-i\mathbf{k}\cdot\mathbf{R}_j} \, p_1(\mathbf{R}_j) = \widetilde{p}_1(\mathbf{k}),$$

the Fourier transform of the single-step PDF. Thus

$$\widetilde{p}(\mathbf{k}, n\tau) = [\widetilde{p}_1(\mathbf{k})]^n \; .$$

Next, we must find the single-step PDF $p_1(\mathbf{r})$ and calculate its Fourier transform. The only condition imposed on a step is that its magnitude be equal to l. Therefore $p_1(\mathbf{r})$ must be proportional to $\delta(r - l)$. But normalization implies that $\int d^3r \, p_1(\mathbf{r})$ must be equal to unity. This condition fixes the constant of proportionality. The normalized PDF is

$$p_1(\mathbf{r}) = \frac{1}{4\pi l^2} \, \delta(r - l).$$

Therefore

$$\widetilde{p}_1(\mathbf{k}) = \int d^3r \, e^{-i\mathbf{k}\cdot\mathbf{r}} \, p_1(\mathbf{r}) = \frac{1}{4\pi l^2} \int d^3r \, e^{-i\mathbf{k}\cdot\mathbf{r}} \, \delta(r - l).$$

Carrying out the integration, we get (check this out)

$$\widetilde{p}_1(\mathbf{k}) = \frac{\sin \, kl}{kl} \, , \;\; \text{where } k = |\mathbf{k}|.$$

The characteristic function of the position vector $\mathbf{r}$ for an n-step random flight is therefore given by

$$\widetilde{p}(\mathbf{k}, n\tau) = \Big(\frac{\sin\, kl}{kl}\Big)^n.$$

Hence

$$p(\mathbf{r}, n\tau) = \frac{1}{(2\pi)^3} \int d^3k\; e^{i\mathbf{k}\cdot\mathbf{r}}\; \Big(\frac{\sin\, kl}{kl}\Big)^n.$$

This is an *exact* formula for the PDF of the end-to-end displacement vector $\mathbf{r}$ in a random flight of n steps, each of length l, in three-dimensional space. Observe its striking similarity to the expression found earlier for the PDF of a sum of n random variables uniformly distributed in $[0, 1]$—a problem that is seemingly quite different from the problem of random flights!

The integration over all the directions of $\mathbf{k}$ is easily carried out if we choose spherical polar coordinates with the polar axis along the direction of $\mathbf{r}$, exploiting rotational invariance. The result is

$$p(\mathbf{r}, n\tau) = \frac{1}{2\pi^2 r} \int_0^\infty dk\, k\, \sin\, kr \Big(\frac{\sin\, kl}{kl}\Big)^n.$$

The evaluation of the integral above for successive values of n is an interesting exercise in its own right. But our present purpose is to examine $p(\mathbf{r}, n\tau)$ for very *large* n—in fact, in the limit $n \to \infty\,,\ \tau \to 0$ such that $\lim\, n\tau = t$. Now, $|(\sin\, kl)/(kl)| < 1$ for all $kl \neq 0$. Hence the factor $[(\sin\, kl)/(kl)]^n$ in the integrand causes the integral to vanish as $n \to \infty$ as long as l remains finite and non-zero. But if l tends to zero simultaneously, we have

$$\Big(\frac{\sin\, kl}{kl}\Big)^n \simeq \Big(1 - \frac{k^2\, l^2}{6}\Big)^n.$$

It is then clear that the right-hand side tends to a finite nontrivial limit (that is not 0 or 1) if, and only if, $l^2 \to 0$ like n^{-1}. This is the *only* possible way in which a nontrivial limiting PDF can arise. But $n^{-1} \sim \tau$, so that we must let l^2 tend to zero like τ. Therefore, let $\tau \to 0$ and $l \to 0$, such that

$$\lim_{l,\,\tau \to 0}\; \frac{l^2}{6\tau} = D,$$

where D is a finite positive quantity called the **diffusion coefficient**. The choice of the precise numerical factor $\frac{1}{6}$ in the above is unimportant. It has been tailored so as to obtain the same numerical factors in the limiting expression for $p(\mathbf{r}, n\tau)$ as the fundamental Gaussian solution of the diffusion equation. In this 'diffusion limit', we have

$$\Big(1 - \frac{k^2\, l^2}{6}\Big)^n = \Big(1 - \frac{Dk^2 t}{n}\Big)^n \underset{n\to\infty}{\longrightarrow} e^{-Dk^2 t}.$$

We thus obtain

$$p(\mathbf{r}, n\tau) \to p(\mathbf{r}, t) = \frac{1}{(2\pi)^3} \int d^3k \; e^{-Dk^2 t + i\mathbf{k}\cdot\mathbf{r}} \, .$$

This integral is most easily evaluated in *Cartesian* coordinates—it then factors into the product of three Gaussian integrals. Put $\mathbf{r} = (x, y, z)$, $\mathbf{k} = (k_1 \, , \, k_2 \, , \, k_3)$, and use the familiar formula for the shifted Gaussian integral. Re-combine $x^2 + y^2 + z^2$ into r^2, to arrive finally at

$$p(\mathbf{r}, t) = \frac{1}{(4\pi D t)^{3/2}} \, e^{-r^2/(4Dt)}.$$

★ **23.** Work through the steps of the derivation to obtain the result quoted above for $p(\mathbf{r}, t)$.

The expression we have found for $p(\mathbf{r}, t)$ is precisely the normalized fundamental Gaussian solution to the three-dimensional **diffusion equation**

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = D \, \nabla^2 p(\mathbf{r}, t)$$

with the initial condition $p(\mathbf{r}, 0) = \delta^{(3)}(\mathbf{0})$ and boundary condition $p(\mathbf{r}, t) \to 0$ as $r \to \infty$. Thus, in the diffusion limit, the random flight goes over into the diffusion process. It is easily checked that the mean squared displacement is

$$\langle r^2 \rangle = \int d^3r \, r^2 \, p(\mathbf{r}, t) = \frac{4\pi}{(4\pi D t)^{3/2}} \int_0^\infty dr \, r^4 \, e^{-r^2/(4Dt)} = 6Dt.$$

The linear growth of the variance of the displacement with time is thus retained in the diffusion limit.

2.12 Stable distributions

As mentioned earlier, a topic of great importance in statistics is the probability distribution of the sums of independent, identically-distributed random variables (*iidrv*). In particular, a crucial question is the existence of a limit law when the number of random variables becomes infinite. We've already seen an example of such a limit law, namely, the Gaussian distribution. The Gaussian is one member (perhaps the most important member, as far as physical applications are concerned) of a whole family of distributions called the **stable distributions**, which are intimately connected with such limit laws.

Suppose we have n *iidrv* $X_1\,,\ X_2\,,\ \ldots\,,\ X_n\,$, each with a cumulative distribution function F. We ask: Are there forms of F such that the sum $\sum_{i=1}^{n} X_i\,$, possibly shifted by an n-dependent constant and re-scaled by another n-dependent constant, also has the *same* distribution function? The complete answer to this question is one of the key results in statistics. There is a whole family of distributions with the property required, called the **stable distributions**. (The full name of the family is **Lévy skew alpha-stable distributions**.) There are several alternative (and equivalent) ways of stating the defining property of these probability distributions.

Definition 1: The CDF F is a stable distribution if and only if, for *every* positive integer $n \geq 2$, it is possible to find a positive constant a_n and a real constant b_n such that the probability distribution of the quantity

$$Z_n = \frac{1}{a_n}\Big\{\sum_{i=1}^{n} X_i - b_n\Big\}$$

is also given by F itself. If this condition can be satisfied with $b_n = 0$ for all n, the distribution F is said to be *strictly* stable. The latter constitute a subset of the class of stable distributions.

Definition 2: Let F be the CDF of two *iidrv* X_1 and $X_2\,$. Then F is a stable distribution if and only if, given any two arbitrary positive numbers a_1 and a_2, a positive number a and a real number b can be found such that $(a_1\,X_1 + a_2\,X_2 - b)/a$ also has the distribution F. Again, if this can be done without a shift constant b, the distribution is strictly stable.

This formulation of the defining property can be re-expressed as a property of the distribution function itself, as follows.

Definition 3: F is a stable distribution if and only if, given any two positive numbers a_1 and $a_2\,$, we can find a positive number a and a real number b such that F satisfies

$$F(x/a_1) * F(x/a_2) = F\left((x-b)/a\right),$$

where the symbol $*$ denotes the convolution of the two distributions. If this relation is satisfied with $b = 0$ in all cases, F is strictly stable.

2.13 The characteristic function of stable distributions

Not surprisingly, the most explicit way of specifying the stable distributions is in terms of their characteristic functions. Definition 3, involving the convolution of distributions, suggests that the characteristic function of a stable distribution might satisfy some sort of 'multiplication property', and be related to exponential

functions. This intuitive guess is indeed borne out.

The family of stable distributions is characterized by four different parameters, but I shall not go into these details here. The important points are the following.

(a) It turns out that the scaling constant a_n in the definition of Z_n in Definition 1 must have the power-law form $a_n = n^{1/\alpha}$, where $0 < \alpha \leq 2$.

(b) The exponent α is the primary characterizer of the members of the family of stable distributions.

(c) The stable distributions have continuous PDFs that are unimodal (i.e., have single peaks).

Let X denote a random variable with a stable distribution, and $p(x)$ its PDF. Except for certain special values of the exponent α, $p(x)$ cannot be written down in explicit closed form, in general. The characteristic function $\widetilde{p}(k)$, however, *can* be expressed in such a form for all $0 < \alpha \leq 2$.

(d) In essence, the modulus $|\widetilde{p}(k)|$ of the characteristic function for a stable distribution with exponent α behaves like $\exp(-|k|^{\alpha})$.

The reasons for restricting the exponent α to the range $0 < \alpha \leq 2$ are as follows:

(i) If $\alpha \leq 0$, the function $\widetilde{p}(k) \to 1$ as $|k| \to \infty$. Therefore $\int_{-\infty}^{\infty} dk\, e^{ikx}\, \widetilde{p}(k)$ diverges, and hence the PDF $p(x)$ does not exist. (The trivial case when $\widetilde{p}(k)$ is a constant corresponds to a δ-function form for $p(x)$.)

(ii) At the other end, if $\alpha > 2$, the inverse Fourier transform of $\widetilde{p}(k)$ is no longer guaranteed to be a real, non-negative function of x, as it must be in order to represent a PDF. (This is much harder to prove.)

2.14 The three important special cases

There are three notable and important cases in which the formula for $\widetilde{p}(k)$ can be inverted to yield explicit expressions for the PDF $p(x)$ in terms of *elementary* functions. There do exist other cases in which the PDF can be obtained explicitly in terms of known functions. But these do not occur as frequently in physical applications. Moreover, even in these cases $p(x)$ is only expressible in terms of certain hypergeometric functions, and does not reduce to any elementary function. The three cases that follow are also the cases that appear most often in physical applications.

(i) The **Gaussian distribution**: For $\alpha = 2$, we have the Gaussian distribution parametrized by the mean μ and variance σ^2. Recall that the PDF and the characteristic function are, respectively,

$$p(x) = (2\pi\sigma^2)^{-1/2}\, e^{-(x-\mu)^2/2\sigma^2} \quad \text{and} \quad \widetilde{p}(k) = e^{-i\mu\, k - \sigma^2\, k^2/2},$$

where $x, k \in (-\infty, \infty)$.

(ii) The **Cauchy distribution**: For $\alpha = 1$, we have the Cauchy distribution. The PDF and characteristic function are given, respectively, by

$$p(x) = (\lambda/\pi)[(x-\mu)^2 + \lambda^2]^{-1} \quad \text{and} \quad \widetilde{p}(k) = e^{-i\mu\, k - \lambda\, |k|},$$

where $x, k \in (-\infty, \infty)$, and $\lambda(> 0), \mu$ are real parameters. The PDF has a Lorentzian shape, symmetric about its center μ. (This is the simplest form of the Cauchy distribution, and corresponds to special values of the parameters other than α that characterize a stable distribution.)

(iii) The **Lévy distribution**: For $\alpha = \frac{1}{2}$, we have the Lévy distribution, given by the PDF

$$p(x) = [c/(2\pi x^3)]^{1/2}\, e^{-c/(2x)} \quad \text{and} \quad \widetilde{p}(k) = e^{-c|k|^{1/2}\,(1 + i\,\mathrm{sgn}\,k)},$$

where c is a positive parameter. Note that $x \in [0, \infty)$ and $k \in (-\infty, \infty)$ in this case. The PDFs in cases (ii) and (iii) actually correspond to the most important cases of more general stable distributions with $\alpha = 1$ and $\alpha = \frac{1}{2}$, respectively, for some special values of certain other parameters.

The Gaussian distribution, corresponding to $\alpha = 2$, differs in a crucial respect from all the other stable distributions.

(i) The stable distributions with $\alpha < 2$ are **heavy-tailed distributions**, in the sense that the PDF has a leading asymptotic behavior $p(x) \sim |x|^{-(\alpha+1)}$ as $|x| \to \infty$. The Gaussian does not share this property.

(ii) As a consequence, *of all the stable distributions, only the Gaussian has a finite variance.* All the other stable distributions have infinite variance.

(iii) For $\alpha \leq 1$, even the first moment or mean value does not exist. Once again, this happens because the PDF $p(x)$ decays to zero too slowly as $|x| \to \infty$.

Finally, I mention that there is a generalization of the Central Limit Theorem that is applicable to the family of stable distributions.

2.15 Connections between the three cases

There are close relationships among the three special distributions listed above. Here are a couple of these.

★ **24.** We've seen that the square of a Gaussian random variable (e.g., the kinetic energy of a molecule in a classical ideal gas) has a gamma distribution. Now consider the *reciprocal* of the square. Let X be a Gaussian random variable with zero mean and variance σ^2. Consider the random variable $\xi = 1/X^2$. Its PDF is given by

$$\rho(\xi) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} dx\, e^{-x^2/(2\sigma^2)}\, \delta(\xi - x^{-2}).$$

Convert the δ-function to δ-functions at $x = \pm\xi^{-1/2}$, and carry out the integration over x. Show that the result is

$$\rho(\xi) = (2\pi\sigma^2\xi^3)^{-1/2}\, e^{-1/(2\sigma^2\xi)} \quad (0 \le \xi < \infty),$$

which is a Lévy distribution with $c = 1/\sigma^2$.

An interesting duality: The result just derived is a special case of a **duality** that exists between different stable distributions:

- A stable distribution with exponent α (where $1 \le \alpha \le 2$) for the random variable X is essentially equivalent to a stable distribution with exponent α^{-1} (so that $\frac{1}{2} \le \alpha^{-1} \le 1$) for the random variable $X^{-\alpha}$.

★ **25.** The *ratio* of two Gaussian random variables, each with a mean equal to zero, is Cauchy-distributed. Let's see how this happens, using as an example the physical context of two independent random walks on a line, in the diffusion limit. Consider two particles, each starting from the origin at $t = 0$, diffusing on the x-axis. The particles are assumed to have no interaction with each other (and to be able to 'pass through' each other). Let X_1 and X_2 be the positions of the particles, and let D_1 and D_2 be their respective diffusion coefficients. The respective PDFs $p_1(x_1\,,\,t)$ and $p_2(x_2\,,\,t)$ of X_1 and X_2 are given by the fundamental Gaussian solution to the one-dimensional diffusion equation, namely,

$$p_i(x_i\,,\,t) = \frac{1}{(4\pi D_i t)^{1/2}}\, e^{-x_i^2/(4D_i t)}, \quad \text{where } i = 1,\, 2.$$

Let $\xi = X_1/X_2$ be the ratio of the positions of the two particles. It is clear that the sample space of ξ is also $(-\infty, \infty)$. The PDF of ξ is given by

$$\rho(\xi\,,\,t) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\; p_1(x_1\,,\,t)\, p_2(x_2\,,\,t)\, \delta\Big(\xi - \frac{x_1}{x_2}\Big)\,.$$

Carry out the integrations to obtain

$$\rho(\xi\,,\,t) = \frac{\lambda}{\pi}\,\frac{1}{(\xi^2+\lambda^2)}\,,\quad \text{where } \lambda = \sqrt{D_2/D_1}.$$

Note that there is no time-dependence on the right-hand side! Remarkably enough, the ratio $\xi = X_1/X_2$ has precisely the same Cauchy distribution for *all* $t > 0$.

2.16 Infinitely divisible distributions

To recapitulate: Let $X_1\,,\,\ldots\,,\,X_n$ be a set of *iidrv* with a stable probability distribution. Then, there exist constants a_n and b_n such that the (shifted and re-scaled) sum $Z_n = \big(\sum_{i=1}^{n} X_i - b_n\big)/a_n$ also has the same stable distribution, for *every* positive integer n.

Going the other way, it is natural to ask the opposite question: *given* a random variable X with a specified probability distribution, when can we write it as the sum of n *iidrv* $X_1\,,\,\ldots\,,\,X_n$ for *every* positive integer value of n? Whenever this can be done, X is said to be an *infinitely divisible* random variable. Its probability distribution is an **infinitely divisible distribution**. It should come as no surprise that stable distributions are intimately related to infinitely divisible distributions.

In order to understand the concept of infinite divisibility clearly, the following points must be noted. It is convenient to call the individual X_i (into which X is decomposed) the 'components' of X.

(a) The components X_i need not have the same distribution as X itself! All that is required is that, for each value of n, all the components X_i (where $i \leq 1 \leq n$) have the same distribution.

(b) All stable distributions are infinitely divisible. *The converse is not necessarily true.* Stable distributions comprise a subset of infinitely divisible distributions.

(c) The special feature of a stable distribution is that, in this case, the components X_i also have the *same* distribution as the sum X, for every n.

The divisibility of a random variable is a subtle concept. A random variable may be 'n-divisible', but not 'm-divisible', where n and m are different positive integers. Here's a very simple example. Suppose the sample space of a random variable X is the set of integers $\{0, 1, 2, 3\}$. It can then be decomposed into 3 components $X_1\,,\,X_2$ and $X_3\,,$ where each X_i has the sample space $\{0, 1\}$. Thus X has the possibility of being 3-divisible (see below). But it can never be 2-divisible,

because there is no way you can find two *iidrv* X_1 and X_2 that sum up to X with the sample space specified above, namely, $\{0, 1, 2, 3\}$. It's not hard to see why. If X_1 and X_2 have the sample space $\{0, 1\}$, then their sum can never reach the value 3. If the sample space of X_1 and X_2 is $\{0, 1, 2\}$, then their sum can have the value 4, which is not in the sample space of X. If $\frac{3}{2}$ is in the sample space of X_1 and X_2, then, since 0 must also be in this sample space, $\frac{3}{2}$ must also be in the sample of space of X, which is not the case. And so on.

Continuing with this example, it must also be understood that not every random variable with the sample space $\{0, 1, 2, 3\}$ is even 3-divisible. For, if each component X_i has the sample space $\{0, 1\}$ with probabilities p and $q = 1 - p$ for the two values (and this is the only possibility), then the distribution of X *must necessarily be a binomial distribution* with exponent 3. That is, the respective probabilities of the values $0, 1, 2$ and 3 must be p^3, $3p^2q$, $3pq^2$ and q^3, in order that X be 3-divisible. It should now be obvious that a random variable X with a binomial distribution with parameter N is N-divisible: each of its N components is a Bernoulli trial, i.e., it can take the values 0 and 1, with respective probabilities p and q.

(d) A moment's thought tells us that, if X is an *infinitely* divisible random variable, then its sample space *must* be unbounded in at least one direction—from below or from above, or both.

The definitive property of an infinitely divisible distribution is found by looking at its characteristic function. We know that the characteristic function of a sum of *iidrv* is just the product of the characteristic functions of its components. Hence:

(e) X can be an infinitely divisible random variable only if its characteristic function $\widetilde{p}(k)$ can be written as the n^{th} power of a characteristic function (of any of its components X_i) for every n. That is, for *every* positive integer n, we must be able to find a characteristic function $\widetilde{p}_n(k)$ such that

$$\widetilde{p}(k) = \big[\, \widetilde{p}_n(k) \,\big]^n.$$

Recall, at this stage, the properties required of a characteristic function, listed earlier. The characteristic functions of the all the Lévy skew alpha-stable distributions can be shown to satisfy these properties. Hence all stable distributions are also infinitely divisible distributions, as already stated. It is easy to see how this happens in the special cases of the Gaussian, Cauchy and Lévy distributions. Their characteristic functions have been written down in the foregoing. It is almost trivial to observe that the characteristic functions of the components in these cases are given by

$$\widetilde{p}_n(k) = \begin{cases} \exp\, \big\{-i(\mu/n)\, k - \frac{1}{2}(\sigma/\sqrt{n}\,)^2\, k^2\big\} & \text{(Gaussian)} \\ \exp\, \{-i(\mu/n)\, k - (\lambda/n)\, |k|)\} & \text{(Cauchy)} \\ \exp\, \big\{-(c/n)\, |k|^{1/2}\, (1 + i\, \operatorname{sgn} k)\big\} & \text{(Lévy).} \end{cases}$$

In effect, all that needs to be done is to *scale down the parameters* of the distributions by n (and by $\sqrt{n}$ in the case of the standard deviation of the Gaussian), to read off $\widetilde{p}_n(k)$ from $\widetilde{p}(k)$ itself. It is obvious that the exponential form of the characteristic function makes this special property possible.

2.17 Infinite divisibility does not imply stability

It remains to give instances of distributions that are infinitely divisible, but are not stable distributions. A prominent example is the Poisson distribution, which is not a member of the family of Lévy skew alpha-stable distributions. Recall that the characteristic function of a Poisson distribution with mean value μ is is $\widetilde{p}(k) = \exp\{\mu\,(e^{-ik} - 1)\}$. But this expression can be written as

$$\widetilde{p}(k) = \left[\exp\left\{(\mu/n)\,(e^{-ik} - 1)\right\}\right]^n$$

for every positive integer n, entailing a simple re-scaling of the mean value. It is therefore obvious that the Poisson distribution is infinitely divisible. Moreover, for each n, the components of a Poisson-distributed random variable are themselves Poisson-distributed.

Similarly, the characteristic function of a Skellam distribution with mean value $(\mu - \nu)$ and variance $(\mu + \nu)$ can be written as

$$\widetilde{p}(k) = \left[\exp\left\{(\mu/n)\,(e^{-ik} - 1) + (\nu/n)\,(e^{ik} - 1)\right\}\right]^n$$

where n is any positive integer. Once again, it is evident that the distribution is infinitely divisible, and that each component for any n also has a Skellam distribution.

★ **26.** We've already seen that the binomial distribution with parameter N is N-divisible into N Bernoulli trials (recall that $\mu = Np$ in this case). Now show that the negative binomial distribution with parameter N is N-divisible into geometric distributions. Start with the fact that the characteristic function of the negative binomial distribution with parameter N and mean μ is given by

$$\widetilde{p}(k) = \left(\frac{N}{N + \mu - \mu\, e^{-ik}}\right)^N.$$

3 Stochastic processes

A **stochastic process** or random process is one in which the random variable takes on different values from its sample space as time elapses. Stochastic processes comprise a rather extensive subject by themselves. Our focus will be on an important sub-class of these processes, namely, Markov processes. These are the processes that are used most often in applications. As usual, we deal with probabilities P and probability densities p, respectively, for discrete and continuous sample spaces. For notational simplicity, we shall deal with the discrete case for the most part. But what follows is adaptable, with obvious modifications, to the continuous case.

Consider a random variable X whose sample space consists of a discrete set of N values, $\{x_1\,,\,\ldots\,,\,x_N\}$. Instead of referring to the values themselves, it is notationally convenient to specify them as 'states' labeled from 1 to N. The indices j, k, l, ... will be used to denote these state labels. Alternatively, I will use $j_1\,,\,j_2\,,\,\ldots$ for the state labels, when the need arises to introduce an arbitrarily large number of labels. The formalism below is also applicable when N is infinite, provided certain convergence conditions are satisfied. As the random process evolves in time, the state of the variable changes randomly.

Let $t_1 < t_2 < t_3 < \ldots$ be an arbitrary sequence of instants of time. The statistical properties of the random process are specified completely by an infinite set of multiple-time **joint probabilities**

$$P_1(j,\,t_1),\; P_2(k,\,t_2\,;\,j,\,t_1),\; P_3(l,\,t_3\,;\,k,\,t_2\,;\,j,\,t_1),\;\ldots\quad \textit{ad infinitum}.$$

The n-time joint probability $P_n(j_n\,,\,t_n\,;\,j_{n-1}\,,\,t_{n-1}\,;\,\ldots\,;\,j_1\,,\,t_1)$ is the probability that the random variable has the values corresponding to the state j_1 at time t_1, the state j_2 at time t_2, and so on, and the state j_n at time t_n. This n-time joint probability is expressible as the product of an n-time **conditional probability** and an $(n-1)$-time joint probability, according to the **chain rule**

$$\begin{aligned} P_n(j_n\,,\,t_n\,;\,j_{n-1}\,,\,t_{n-1}\,;\ldots;\,j_1\,,\,t_1) = {}& P_n(j_n\,,\,t_n\,|\,j_{n-1}\,,\,t_{n-1}\,;\,\ldots\,;\,j_1\,,\,t_1)\times \\ & \times P_{n-1}(j_{n-1}\,,\,t_{n-1}\,;\,j_{n-2}\,,\,t_{n-2}\,;\,\cdots\,;\,j_1\,,\,t_1). \end{aligned}$$

The first factor on the right-hand side is a conditional probability: it is the probability that the state j_n occurs at time t_n, *given* that the events to the right of the vertical bar have occurred at the successively earlier instants $t_{n-1}\,,\,t_{n-2}\,,\,\ldots\,,\,t_1$, respectively. In turn, the joint probability P_{n-1} on the right-hand side can be written as the product of an $(n-1)$-time conditional probability and a joint probability P_{n-2}, and so on. Thus, a stochastic process is specified by an infinite hierarchy of multiple-time conditional probabilities, and a single-time probability $P_1(j,t)$. This is convenient: in all physical applications of probability and random

processes, we can only write equations for, or model, *conditional* probabilities or probability densities—rather than the corresponding absolute probabilities themselves.

3.1 Discrete Markov processes

In practice, it is essentially impossible to determine or specify all the members of the infinite set of conditional probabilities. We must make further simplifying assumptions. Fortunately for us, most physical stochastic processes are also amenable to modeling in terms of such simplified processes. The most notable of these is the class of **Markov processes**.

Markov processes can have discrete or continuous sample spaces. Discrete Markov processes may also be defined in discrete *time*, in which case they are called **Markov chains**. There is a vast literature on this subject. But we shall restrict ourselves here to the case of continuous time. This is the case that occurs most commonly in physical applications such as nonequilibrium statistical mechanics. The discussion below is an appropriate extension of what holds good for Markov chains.

A Markov process is one with a 'memory' that is restricted, at any instant of time, to the *immediately preceding time argument alone*. That is,

$$P_n(j_n\,,\,t_n\,|\,j_{n-1}\,,\,t_{n-1}\,;\,\ldots\,;\,j_1\,,\,t_1) = P_2(j_n\,,\,t_n\,|\,j_{n-1}\,,\,t_{n-1}) \quad \text{for all} \quad n \geq 2.$$

The single-time-step memory characterizing a Markov process is equivalent to saying that the *future* state of the process is only dependent on its *present* state, and not on the history of *how* the process reached the present state. Although this appears to be a mild technical assumption, it leads to a great deal of simplification. It has an immediate implication: *All* the joint probabilities of a Markov process are expressible as products of just two independent probabilities:

(i) The *single-time* probability $P_1(j\,,\,t)$.

(ii) The *two-time* conditional probability $P_2(k\,,\,t\,|\,j\,,\,t')$, where $t' < t$.

The Markov assumption immediately simplifies the n-time joint probability to

$$\begin{aligned} P_n(j_n\,,\,t_n\,;\,j_{n-1}\,,\,t_{n-1}\,;\ldots;\,j_1\,,\,t_1) = \quad & P_2(j_n\,,\,t_n\,|\,j_{n-1}\,,\,t_{n-1}) \times \\ & P_2(j_{n-1}\,,\,t_{n-1}\,|\,j_{n-2}\,,\,t_{n-2}) \cdots \times \\ & P_2(j_2\,,\,t_2\,|\,j_1\,,\,t_1)\,P_1(j_1\,,\,t_1). \end{aligned}$$

Further simplification occurs in the case of a **stationary random process**, which is a random process whose *statistical* properties do not depend on the

choice of the origin of time. You will recognize that stationarity is the probabilistic equivalent of the assumption of **time-translation invariance** in deterministic dynamics (classical or quantum). In the mathematics literature, a stationary random process is often called a *homogeneous* random process, because stationarity is the same thing as homogeneity in time. For a stationary random process, the single-time probability $P_1(j,t)$ is actually *independent* of t, i.e.,

$$P_1(j\,,\,t) = P_1(j) \quad \text{(stationary process)},$$

and the two-time conditional probability is a function of the time *difference* $(t-t')$ alone, i.e.,

$$P_2(k\,,\,t\,|\,j\,,\,t') = P_2(k\,,\,t-t'\,|\,j,0) \quad \text{(stationary process)}.$$

It is convenient to use the notation

$$P_2(k\,,\,t-t'\,|\,j) \stackrel{\text{def.}}{=} P_2(k\,,\,t-t'\,|\,j,0).$$

Further, let's drop the subscripts 1 and 2 in P_1 and P_2, and write the two different functions as simply $P(j)$ and $P(k\,,\,t-t'\,|\,j)$, for notational simplicity. No confusion should arise, because the former probability is time-independent, while the latter is not. Whenever necessary, I shall refer to $P(j)$ as the *stationary probability*, and to $P(k\,,\,t-t'\,|\,j)$ or $P(k\,,\,t\,|\,j)$ as the *conditional probability*. We therefore have, for a stationary Markov process,

$$P_n(j_n\,,\,t_n\,;\,j_{n-1}\,,\,t_{n-1}\,;\,\ldots\,;\,j_1\,,\,t_1) = \left\{\prod_{r=1}^{n-1} P(j_{r+1}\,,\,t_{r+1}-t_r\,|\,j_r)\right\}\,P(j_1),$$

for every $n \geq 2$. From this point onward, I consider stationary processes, unless otherwise specified. This is entirely for convenience—the notation simplifies somewhat in the case of stationary processes.

Next, we need inputs for the stationary probability $P(j)$ and the conditional probability $P(k\,,\,t\,|\,j)$. These are, *a priori*, independent quantities. Now, in physical situations modeled by Markov processes, the dynamics underlying the random process generally enjoys a sufficient degree of **mixing**. This is a technical term in dynamical systems theory. It refers to a kind of 'scrambling up' that occurs in phase space under time evolution for most real-life dynamical systems. (I will not digress into further detail here.) The mixing property ensures that the following important relationship holds good:

$$\lim_{t\to\infty} P(k\,,\,t\,|\,j) = P(k).$$

In other words, the memory of the initial state j is 'lost' as $t \to \infty$, and the conditional probability simply tends to the stationary probability $P(k)$ corresponding to the final state k. As a consequence, *the single conditional probability*

distribution $P(k\,,\,t\,|\,j)$ *completely determines* all *the statistical properties of such a stationary Markov process.*

The property implied by the relationship above appears to be quite plausible on physical grounds. But there are some technical issues involved here. One of them, which is pertinent to current work in nonequilibrium statistical physics, is whether there exists a *unique* stationary distribution, as opposed to a whole set of such distributions. The latter possibility is indeed realized in many physical situations.

3.2 The master equation

The next step is to find an equation satisfied by the conditional probability $P(k\,,\,t\,|\,j)$ of a Markov process. A starting point for the derivation of this equation is the **Chapman-Kolmogorov equation** satisfied by Markov processes:

$$P(k\,,\,t\,|\,j) = \sum_{l=1}^{N} P(k\,,\,t-t'\,|\,l)\;P(l\,,\,t'\,|\,j) \quad \text{for any } \; t' \in (0,t).$$

In effect, this equation says that the probability of 'propagating' from an initial state j to a final state k is the product of the probabilities of propagating from j to any *intermediate* state l, and subsequently from l to k, summed over all possible intermediate states. The Chapman-Kolmogorov equation is applicable to continuous Markov processes as well. We'll write down the corresponding version of the equation later on.

The Chapman-Kolmogorov equation is a *nonlinear* equation for the conditional probability, since the right-hand side is quadratic in P. Nonlinear equations are generally much harder to handle than linear ones. But in many cases the Chapman-Kolmogorov equation can be reduced to a linear equation, depending on the existence of a transition probability per unit time, or **transition rate** $w(k\,|\,j)$ between any two distinct states j and k, defined as follows. Recall that, for a stationary process, $P(k, t+\delta t\,|\,j, t) = P(k, \delta t\,|\,j)$. We now assume that, in any infinitesimal time interval δt, the probability $P(k, \delta t\,|\,j)$ is proportional to δt, with a coefficient of proportionality that depends on j and k, i.e.,

$$P(k, \delta t\,|\,j) = w(k\,|\,j)\,\delta t \quad (j \neq k).$$

Such a relation is not always guaranteed for all Markov processes. When it does hold good, it *defines* the transition rate $w(k\,|\,j)$. What follows below is therefore valid for those Markov processes for which finite transition rates exist between the states. The quantity $w(k\,|\,j)$ is the probability per unit time that the random variable jumps from (its value in) the state j to (its value in) the state k. It can

then be shown that the Chapman-Kolmogorov equation leads to the following differential equation for the set of conditional probabilities $\{P(k,t\,|\,j)\}$:

$$\frac{d}{dt}P(k,t\,|\,j) = \sum_{\substack{l=1\\ l\neq k}}^{N} \Big\{ w(k\,|\,l)\,P(l,t\,|\,j) - w(l\,|\,k)\,P(k,t\,|\,j) \Big\}.$$

For each initial state j, this equation is satisfied for every k from 1 to N. The initial condition is of course

$$P(k,0\,|\,j) = \delta_{jk}\,.$$

The differential equation above for the conditional probability is called the **master equation** (for a discrete Markov process). It has the form of a **rate equation** of the sort that is familiar, for instance, in chemical physics. Viewed this way, it is clear that the first term on the right is a 'gain' term, while the second is a 'loss' term.

★ **27.** Derive the master equation from the Chapman-Kolmogorov equation.

Hint: Subtract $P(k,t-t'\,|j)$ from both sides of the Chapman-Kolmogorov equation, and set $t' = t - \delta t$. Use the fact that, for *any* given initial state j, the sum over the final states of the conditional probability must be equal to unity, by the conservation of probability. You will need to make use of this relation in the form

$$1 - P(k,\delta t\,|\,k) = \sum_{\substack{l=1\\ l\neq k}}^{N} w(l\,|\,k)\,\delta t.$$

Why are Markov processes so important? Observe that the master equation for the conditional probability distribution is a *first-order* differential equation in the time variable. *This is a direct consequence of the Markovian nature of the random process.* The solution of the equation requires a single piece of information: namely, the specification of the initial distribution $P(k,0\,|\,j)$. No *earlier* history is necessary, i.e., no information is required about *how* that initial distribution was attained. But this is precisely what holds good in the case of *deterministic* dynamics as well, both classical and quantum! In classical dynamics, Hamilton's equations for the variables in phase space are first-order differential equations in t. In quantum mechanics, the Schrödinger equation for the state vector, or the Liouville equation for the density matrix, are first-order differential equations in t. In this sense, the master equation for a Markov process is the stochastic analog of the fundamental equations of deterministic dynamics.

Recall that classical dynamics is 'first-order' dynamics, provided we identify and work in terms of the right number of dependent variables, i.e., the generalized

coordinates *as well as* the generalized momenta pertaining to a system. (A simple example: as you know, underlying the familiar second-order equation $m\ddot{x} = F(x)$ is the pair of coupled *first*-order equations $\dot{x} = p/m$ and $\dot{p} = F(x)$.) Similarly, a given physical situation may appear to require modeling by a stochastic process that is more history-dependent than a Markov process. But it often turns out that this is in fact a Markov process in a coupled *set* of random variables, i.e., by a *multi-component* Markov process. This possibility is the reason why Markov processes make their appearance everywhere.

3.3 Formal solution of the master equation

Next, we turn to the formal solution of the master equation. Let $\mathsf{P}(t)$ denote the column vector whose k^{th} row is $P(k, t \,|\, j)$. Although the initial state j is not explicit in this notation, we'll keep it in mind. The master equation can then be written as the matrix equation

$$\frac{d\mathsf{P}}{dt} = W\,\mathsf{P},$$

where W is called the **relaxation matrix** in the physics literature. The reason will become clear shortly. This $(N \times N)$ matrix has off-diagonal elements

$$W_{kj} \stackrel{\text{def.}}{=} w(k \,|\, j) \quad (k \neq j)$$

and diagonal elements

$$W_{kk} = -\sum_{\substack{l=1 \\ l \neq k}}^{N} w(l \,|\, k).$$

W is *not* a symmetric matrix, in general.

Since the elements of W do not depend on time (we are considering a stationary process), the formal solution to the matrix differential equation for P is simply

$$\mathsf{P}(t) = e^{Wt}\,\mathsf{P}(0),$$

where $\mathsf{P}(0)$ is a column vector whose j^{th} row is 1, and all other elements are 0. (This is where the initial state j makes its appearance.) It follows that the time-dependence of the probability distribution is essentially determined by the eigenvalue spectrum of the matrix W.

Finding the exponential of the matrix Wt in explicit form is generally not a simple task. As you know, an arbitrary (2×2) matrix is easily exponentiated using the Pauli matrices, but no such simple general formula of this sort exists for matrices of higher order. A more convenient way of dealing with the problem

is to work with Laplace transforms. Let $\widetilde{\mathsf{P}}(s)$ denote the Laplace transform of $\mathsf{P}(t)$. Then

$$\widetilde{\mathsf{P}}(s) = \big(sI - W\big)^{-1}\,\mathsf{P}(0),$$

where I is the $(N \times N)$ unit matrix. The matrix $(sI - W)^{-1}$ is, of course, the **resolvent** of the transition matrix W. The problem of exponentiating the matrix W is replaced by the relatively simpler one of finding the *inverse* of a related matrix. We are still left with the task of inverting the Laplace transform, of course.

The known properties of the relaxation matrix W make it possible to say something about its eigenvalue spectrum. I restrict myself here to the most general statements, without going into the details of the necessary and sufficient conditions for their validity, exceptional cases, and so on. Note that the sum of the elements of each column of W is equal to zero. It follows at once that $\det W = 0$, so that 0 is an eigenvalue of this matrix. (The physical implication of this fact will become clear shortly.) The elements of W are real. Hence its eigenvalues occur in complex conjugate pairs. Now apply **Gershgorin's Circle Theorem** to W. The centers of the Gershgorin discs are located on the negative real axis of the complex plane, at the points $W_{kk} = -|W_{kk}|$, where k runs from 1 to N. The radii of the circles are given by $|W_{kk}|$. The right-most points of all the Gershgorin discs are therefore at the origin. Hence the real parts of all the eigenvalues are negative, except for the eigenvalue 0. It follows that the time evolution of the probabilities, which is governed by e^{Wt}, is given by *decaying* exponentials in t, possibly multiplied by factors involving sinusoidal functions of t. This property justifies the term 'relaxation matrix' for W.

3.4 The stationary distribution

To what distribution, if any, do the conditional probabilities $P(k, t \,|\, j)$ 'relax' as $t \to \infty$? An important aspect of the master equation concerns the conditions under which there exists a unique time-independent distribution $\mathsf{P}^{\rm st}$ whose $k^{\rm th}$ row is the stationary probability $P(k)$. This question becomes even more non-trivial for discrete Markov processes with an an infinite-dimensional state space $(N \to \infty)$, and for continuous Markov processes. Here we shall only consider the case when N is finite and, further, a unique stationary distribution $\mathsf{P}^{\rm st}$ exists. This distribution is given by the solution of

$$\frac{d\mathsf{P}^{\rm st}}{dt} = W\,\mathsf{P}^{\rm st} = 0.$$

The stationary distribution $\mathsf{P}^{\rm st}$ is therefore given by the elements of the appropriately normalized *right* eigenvector of the matrix W corresponding to the eigenvalue 0. As each column of W adds up to zero, it is evident that the uniform *row* vector $\begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix}$ is the (unnormalized) *left* eigenvector of W with eigenvalue

zero. Since W is not symmetric in general, the right and left eigenvectors are not adjoints of each other. It is also necessary to show that the elements of the right eigenvector are non-negative numbers that add up to unity, after normalization. I do not go into the proof of this assertion here, but merely mention that it is based on the **Frobenius-Perron Theorem** for real non-negative matrices.

From the explicit form of the master equation in terms of the transition rates, it is clear that the stationary distribution is given by the solution of the set of N homogeneous simultaneous equations

$$\sum_{\substack{l=1 \\ l\neq k}}^{N} \Big\{ w(k\,|\,l)\,P(l) - w(l\,|\,k)\,P(k) \Big\} = 0, \quad (k = 1, 2, \ldots, N).$$

Since $\det W = 0$, we are guaranteed that this set of homogeneous simultaneous equations has unique nontrivial solutions for the *ratios* of the unknowns, say $P(2)/P(1)$, $P(3)/P(1), \ldots, P(N)/P(1)$. We need one more equation, an *inhomogeneous* one, to fix the actual values of the probabilities. This is provided by the normalization condition $\sum_{k=1}^{N} P(k) = 1$. All the probabilities $\{P(k)\}$ are now fully determined. The stationary distribution $\mathsf{P}^{\rm st}$ is thus expressible in terms of the set of transition rates of the Markov process.

★ **28.** When $N = 2$, we have a 2-state Markov process, also called a **dichotomous Markov process** (DMP). We'll consider the DMP in greater detail shortly. Show that the stationary distribution is given in this case by

$$P(1) = \frac{W_{12}}{W_{12} + W_{21}} \quad \text{and} \quad P(2) = \frac{W_{21}}{W_{12} + W_{21}}\,.$$

★ **29.** The stationary distribution for a general 3-state Markov process is already considerably more complicated. Show that it is given by $P(k) = \mathcal{N}_k/\mathcal{D}_k$, where

$$\mathcal{N}_1 = W_{12}W_{13} + W_{13}W_{32} + W_{12}W_{23}\,,$$
$$\mathcal{D}_1 = \mathcal{N}_1 + W_{21}W_{13} + W_{31}W_{12} + W_{32}W_{21} + W_{23}W_{31} + W_{31}W_{32} + W_{21}W_{23}\,,$$

and $\mathcal{N}_2$, $\mathcal{D}_2$, $\mathcal{N}_3$, $\mathcal{D}_3$ are given by cyclic permutations of the expressions above.

3.5 Detailed balance

It is clear that, even for the stationary distribution, the expressions rapidly get more and more complicated as N increases. But there is a very important special case in which the solution is simplified considerably. This is the situation in which the **detailed balance** condition applies: namely, when *each* term in the summand in the sum representing $W\,\mathsf{P}^{\rm st}$ vanishes by itself. We then have

$$w(k\,|\,l)\,P(l) = w(l\,|\,k)\,P(k) \quad \text{(detailed balance)}$$

for *every pair* of distinct states k and l. The detailed balance condition has its origin in **time reversal invariance**. Its importance arises from the fact that *it is applicable to systems in the state of thermal equilibrium*, under normal circumstances. The corresponding stationary distribution may then be termed the **equilibrium distribution**, with probabilities denoted by $P^{\rm eq}(k)$. These probabilities can be found easily in terms of the transition rates.

★ **30.** Use detailed balance and the normalization condition $\sum_{k=1}^{N} P^{\rm eq}(k) = 1$ to show that

$$P^{\rm eq}(k) = \left\{ 1 + \sum_{\substack{l=1 \\ l \neq k}}^{N} \frac{w(l \,|\, k)}{w(k \,|\, l)} \right\}^{-1} .$$

The case of a symmetric relaxation matrix W**:** In the special case of a symmetric relaxation matrix W, i.e., when $w(l \,|\, k) = w(k \,|\, l)$, more can be said on general grounds. As W is a real symmetric matrix, it is diagonalizable by an orthogonal transformation. All its eigenvalues are real. Taken together with the conclusions drawn earlier based on the Gershgorin Circle Theorem, this implies that the eigenvalues are negative, except for the eigenvalue 0. The corresponding normalized right eigenvector is the equilibrium distribution, which reduces in this case to the *uniform distribution* $P^{\rm eq}(k) = 1/N$ for every k. This should remind you of the fundamental postulate of equilibrium statistical mechanics: all the accessible microstates of an isolated macroscopic system in thermal equilibrium have equal probabilities.

3.6 The autocorrelation function

One of the most important quantities associated with any stochastic process is its **autocorrelation function**, which is a generalization of the variance of a random variable. It is the primary measure of the degree of 'memory' possessed by the random variable as it evolves in time. The expressions given below for the autocorrelation function pertain to the case of a discrete-valued random process $X(t)$. Their extension to a continuous random process is straightforward. These expressions are valid for all stationary discrete stochastic processes, and not just Markov processes. In the case of a Markov process, however, a knowledge of the autocorrelation function is of particular significance. This is because the autocorrelation function depends on the conditional probability, and (as you know) the latter essentially determines *all* the joint probabilities associated with a Markov process.

Consider the average value of the product of the random variable at time t_1

with itself at time t_2. We have, by definition,

$$\begin{aligned}\langle X(t_1)\,X(t_2)\rangle &\stackrel{\text{def.}}{=} \sum_j \sum_k x_j\,x_k\;P_2(k\,,t_2\,;\,j\,,\,t_1)\\ &= \sum_j \sum_k x_j\,x_k\;P_2(k\,,t_2\,|\,j\,,\,t_1)\,P_1(j\,,\,t_1).\end{aligned}$$

For a stationary random process this becomes, on writing $t_2 - t_1$ as simply t,

$$C_X(t) \stackrel{\text{def.}}{=} \langle X(0)\,X(t)\rangle = \sum_j \sum_k x_j\,x_k\;P(k\,,t\,|\,j)\,P(j),$$

in terms of the stationary probability and the conditional probability. This is the autocorrelation function $C_X(t)$ of the stationary, discrete random process $X(t)$, in the case when the mean value is zero at all times.

When the mean value of a stationary random process is nonzero, its autocorrelation function is defined in terms of the *deviation* from the mean value, namely, $\delta X = X - \langle X\rangle$, according to $C_X(t) \stackrel{\text{def.}}{=} \langle \delta X(0)\,\delta X(t)\rangle$.

★ **31.** Show that

$$C_X(t) = \Big\{ \sum_j \sum_k x_j\,x_k\;P(k\,,t\,|\,j)\,P(j)\Big\} - \langle X\rangle^2,$$

where $\langle X\rangle = \sum_j x_j\,P(j)$.

3.7 The dichotomous Markov process

The dichotomous Markov process (or DMP) is a very basic random process that appears in various forms in numerous applications in science and engineering. It is a stationary Markov, discrete (or jump) process $X(t)$ in which the sample space consists of just two values, c_1 and c_2. The random variable X flips back and forth between these values (or between the states 1 and 2) at random instants of time. Let λ_1 be the mean transition rate from c_1 to c_2, and λ_2 the mean transition rate from c_2 to c_1. Successive transitions are supposed to be completely uncorrelated with each other. The transition matrix for this 2-state process is therefore

$$W = \begin{pmatrix} -\lambda_1 & \lambda_2 \\ \lambda_1 & -\lambda_2 \end{pmatrix}.$$

Before going on to the solution of the master equation in this case, consider the physical significance of the parameters λ_1 and λ_2. Suppose the system is in the state 1 at any instant of time. The probability that it will make a transition

to the state 2 in the next infinitesimal time interval δt is just $\lambda_1\,\delta t$. Now, δt is taken to be so small that the probability of two or more transitions occurring within this time interval is negligible. Hence the probability that it will *not* make a transition in the interval δt is $(1-\lambda_1\,\delta t)$. From this fact, it is easy to calculate the probability that, if the system has just arrived at the state 1 at some instant of time t_0, it remains in that state at time $t\,(\geq t_0)$ without having made any transitions in between. All you have to do is to divide the interval $(t-t_0)$ into n sub-intervals, each of duration δt. Since the process is Markovian, there is no history-dependence. The no-transition or zero-transition probability we seek is therefore $(1-\lambda_1\,\delta t)^n$, in the limit $n\to\infty$ and $\delta t\to 0$ such that the product $n\,\delta t=(t-t_0)$. Hence

$$\begin{aligned}\Pr\big\{\text{stay in state 1 during an interval } (t-t_0)\big\} &= \lim_{\substack{n\to\infty\\ \delta t\to 0}}(1-\lambda_1\,\delta t)^n\\ &= e^{-\lambda_1(t-t_0)}.\end{aligned}$$

Exactly the same argument applies to the state 2, with λ_1 replaced by λ_2. Thus

$$\Pr\big\{\text{stay in state 2 during an interval } (t-t_0)\big\}=e^{-\lambda_2(t-t_0)}.$$

It follows that the **mean residence time** in the state 1 between two successive transitions is $\tau_1=\lambda_1^{-1}$. Similarly, the mean residence time in state 2, between successive transitions, is $\tau_2=\lambda_2^{-1}$. We may therefore expect that, over a very long interval of time, the fraction of the total time that the system spends in states 1 and 2 are, on the average,

$$\frac{\tau_1}{\tau_1+\tau_2}=\frac{\lambda_2}{\lambda_1+\lambda_2}\quad\text{and}\quad\frac{\tau_2}{\tau_1+\tau_2}=\frac{\lambda_1}{\lambda_1+\lambda_2}.$$

This would imply that the *a priori* probabilities $P(1)$ and $P(2)$ of finding the system in states 1 and 2 are, respectively, precisely the fractions above. But you have already seen that this is indeed so—recall the expressions written down earlier for the stationary probabilities $P(1)$ and $P(2)$ of a 2-state stationary Markov process.

★ **32.** Given these expressions for the stationary probabilities $P(1)$ and $P(2)$, it is easy to write down the mean and variance of X. Show that

$$\langle X\rangle=\frac{\tau_1\,c_1+\tau_2\,c_2}{\tau_1+\tau_2}=\frac{c_1\,\lambda_2+c_2\,\lambda_1}{\lambda_1+\lambda_2}\,,$$

$$\big\langle\big(X-\langle X\rangle\big)^2\big\rangle=\frac{\tau_1\,\tau_2\,(c_1-c_2)^2}{(\tau_1+\tau_2)^2}=\frac{\lambda_1\,\lambda_2\,(c_1-c_2)^2}{(\lambda_1+\lambda_2)^2}\,.$$

3.8 Solution of the master equation

As already defined for an N-state Markov process, let $\mathsf{P}(t)$ denote the column vector whose k^{th} row is $P(k,t\,|\,j)$. For a given initial state j (which can be either 1 or 2), k now runs over the values 1 and 2. The master equation is $d\mathsf{P}/dt = W\,\mathsf{P}$, where W is the (2×2) matrix given above. The solution is $\mathsf{P}(t) = e^{Wt}\,\mathsf{P}(0)$, where $\mathsf{P}(0) = \binom{1}{0}$ or $\binom{0}{1}$, depending on whether $j = 1$ or 2. The problem therefore reduces to finding the exponential of the matrix Wt.

★ **33.** Let $\lambda \stackrel{\text{def.}}{=} \frac{1}{2}(\lambda_1 + \lambda_2)$ denote the *mean transition rate* for the DMP.

(a) Show that

$$e^{Wt} = \frac{1}{2\lambda}\begin{pmatrix} \lambda_2 + \lambda_1\, e^{-2\lambda t} & \lambda_2\,(1 - e^{-2\lambda t}) \\ \lambda_1\,(1 - e^{-2\lambda t}) & \lambda_1 + \lambda_2\, e^{-2\lambda t} \end{pmatrix}.$$

(b) Hence show that the four normalized conditional probabilities characterizing the DMP are given by

$$\left.\begin{aligned} P(c_1\,,\,t\,|\,c_1) &= \frac{\lambda_2 + \lambda_1\, e^{-2\lambda t}}{\lambda_1 + \lambda_2}\,, \quad P(c_2\,,\,t\,|\,c_1) = \frac{\lambda_1\,(1 - e^{-2\lambda t})}{\lambda_1 + \lambda_2} \\ P(c_1\,,\,t\,|\,c_2) &= \frac{\lambda_2\,(1 - e^{-2\lambda t})}{\lambda_1 + \lambda_2}\,, \quad P(c_2\,,\,t\,|\,c_2) = \frac{\lambda_1 + \lambda_2\, e^{-2\lambda t}}{\lambda_1 + \lambda_2}\,. \end{aligned}\right\}$$

Hint: A simple way to compute e^{Wt} is to note that $W^2 = -2\lambda\, W$.

Observe that

$$\lim_{t\to\infty} e^{Wt} = \frac{1}{2\lambda}\begin{pmatrix} \lambda_2 & \lambda_2 \\ \lambda_1 & \lambda_1 \end{pmatrix}.$$

Hence the stationary probability distribution of the DMP is simply

$$\mathsf{P}^{\text{st}} = \begin{pmatrix} P(1) \\ P(2) \end{pmatrix} = \begin{pmatrix} \lambda_2/(\lambda_1 + \lambda_2) \\ \lambda_1/(\lambda_1 + \lambda_2), \end{pmatrix},$$

as we have found already.

★ **34.** Show that the autocorrelation function of a dichotomous Markov process is given by

$$\langle \delta X(0)\, \delta X(t) \rangle = \frac{\lambda_1\, \lambda_2\, (c_1 - c_2)^2}{(\lambda_1 + \lambda_2)^2}\, e^{-2\lambda t}\,,$$

where $\delta X(t) \equiv X(t) - \langle X \rangle$. Hence a DMP is *exponentially correlated*, i.e., the autocorrelation function is a single decaying exponential in t. The **correlation time** is

$$\tau_{\text{corr}} = \frac{1}{2\lambda} = \frac{1}{\lambda_1 + \lambda_2} = \frac{\tau_1\, \tau_2}{\tau_1 + \tau_2}\,,$$

rather than just λ^{-1}, as one might guess at first sight. In other words, the correlation time of a dichotomous Markov process is the harmonic mean of the mean residence times in the two states of the process.

The symmetric DMP: When $c_1 = -c_2 = c$ (say), and $\lambda_1 = \lambda_2 = \lambda$, we have a *symmetric* DMP, and all the foregoing expressions simplify considerably. We then have

$$\langle X \rangle = 0, \quad \langle X^2 \rangle = c^2, \quad \langle X(0)\, X(t) \rangle = c^2\, e^{-2\lambda t}.$$

Further,

$$P(\pm c\,,\, t \mid \pm c) = e^{-\lambda t} \cosh\, \lambda t\,, \quad P(\pm c\,,\, t \mid \mp c) = e^{-\lambda t} \sinh\, \lambda t.$$

3.9 The Poisson pulse process and radioactive decay

A **birth-and-death process** is a Markov process in which the random variable takes on integer values, and direct transitions only occur between *neighboring* values (or states) with prescribed mean rates. Common examples include the Poisson pulse process and a simple random walk on a linear lattice.

In the symmetric DMP, suppose we focus on the instants of *time* at which the DMP flips from one of its two values to the other. These instants themselves constitute a random sequence called a **Poisson pulse process** with a mean pulse rate λ (called the *intensity* of the Poisson process in the statistics literature): that is, the probability that exactly r such instants occur in a given time interval t is Poisson-distributed, being given by $(\lambda t)^r\, e^{-\lambda t}/r!$. Let's now see how this comes about.

It is instructive to work in the context of a physical example, such as the decay events in a radioactive isotope containing a sufficiently large number of nuclei. The latter condition ensures that, during the time interval in which we observe the sample, the events can be regarded as an on-going *stationary* random process. Starting at an arbitrary instant of time, the sequence of instants at which a decay takes place comprise a Poisson pulse process. The number of nuclei that have decayed till time t is an integer-valued random variable $r(t)$ that starts at the value zero, and increases in steps of unity at random instants of time.

Consider a sufficiently large sample of a radioactive isotope with decay constant λ. At the level of the individual nucleus, λ is the *mean rate* of decay. The basic assumptions are as follows:

(i) In any sufficiently small time interval δt, only one of two things can happen : either a single decay occurs, with probability $\lambda\, \delta t$, or there is no decay at all, with probability $(1 - \lambda\, \delta t)$.

If δt is not sufficiently small, then of course multiple decays can occur in that interval. The idea is that, for any given finite value of λ, the time increment δt can be chosen to be small enough for the assumption (of either no decay or just one decay) to be valid. The probability that more than one decay occurs in the interval δt is at least of the order of $(\delta t)^2$.

(ii) Successive decays are statistically independent of each other.

That is, the locations on the time axis of the instants when a decay occurs are completely uncorrelated with each other. In technical terms, this means that the number of decays is a Markov process. In the present instance, since the number n of decay events cannot decrease as t increases, we have a Markovian **birth process**.

Let $P(r,\, t+\delta t)$ denote the probability that exactly r nuclei have decayed in a time interval $t+\delta t$. This probability is given by the sum of the contributions from two possibilities:

— *Either* $(r-1)$ decays have occurred till time t, for which the probability is $P(r-1,\, t)$, and one more decay occurs in the incremental interval δt, with probability $\lambda\,\delta t$;

— *or*, all r decays have already occurred in time t, for which the probability is $P(r,\, t)$, and no further decay occurs in the incremental interval δt, with probability $1-\lambda\,\delta t$.

As these two possibilities are *mutually exclusive events*, their respective probabilities add up. We therefore have

$$P(r, t+\delta t) = (\lambda\,\delta t)\,P(r-1, t) + (1-\lambda\,\delta t)\,P(r, t), \quad \text{where} \quad r \geq 1.$$

Move the term $P(r,t)$ to the left-hand side, divide both sides by δt and let $\delta t \to 0$. The outcome is the differential equation

$$\frac{dP(r,t)}{dt} = \lambda\,\big\{P(r-1,t) - P(r,t)\big\}, \quad r \geq 1.$$

This is the master equation for the probability distribution. Note that it is actually a set of coupled differential equations for the infinite set of probabilities $\{P(r,t)\}$, where $r = 1, 2, \ldots$. We need an initial condition for each $P(r,t)$, which is provided by the obvious condition $P(r,0) = \delta_{r,0}$. Hence $P(r,0) = 0$ for every $r \geq 1$, while $P(0,0) = 1$.

The zero-decay probability $P(0,t)$ itself obeys an even simpler equation, namely,

$$P(0, t+\delta t) = (1-\lambda\,\delta t)\,P(0,t).$$

The corresponding differential equation is therefore

$$\frac{dP(0,t)}{dt} = -\lambda\, P(0,t).$$

The solution to this equation, with the initial condition $P(0,0) = 1$, is just

$$P(0,t) = e^{-\lambda t}\,.$$

In other words: The probability that no decay event occurs in a time interval t decreases exponentially with increasing t.

The coupled set of equations for $P(r,t)$ can be solved in several ways. As you would expect, a convenient way is to use a generating function. Accordingly, let

$$f(z,t) = \sum_{r=0}^{\infty} P(r,t)\, z^r,$$

where z is a complex variable. Multiply both sides of the equation for $dP(r,t)/dt$ by z^r, sum over r, and add on the equation for $dP(0,t)/dt$. The result is the differential equation

$$\frac{\partial f}{\partial t} = \lambda(z-1)\, f.$$

Since $P(r,0) = \delta_{r,0}$, the initial condition on $f(z,t)$ is $f(z,0) = 1$. The solution for $f(z,t)$ is

$$f(z,t) = e^{\lambda(z-1)t}.$$

It is now trivial to pick out the coefficient of z^r in $f(z,t)$ to get the probability distribution

$$P(r,t) = \frac{e^{-\lambda t}\,(\lambda t)^r}{r!}\,, \quad r \geq 0.$$

Therefore $P(r,t)$ is a Poisson distribution with a mean value λt: that is, the mean number of decays that take place in a time interval t is just λt. This is in complete accord with the identification of λ as the mean rate of decay of a nucleus. The half-life of the isotope concerned is $(\ln 2)/\lambda$.

3.10 Biased random walk on a linear lattice

Next, let's consider the prototype of a **birth-and-death process**, namely, a biased random walk on an infinite linear lattice in continuous time. The sites of the lattice are labelled by the integer j. The random walker jumps from any site j to the site $(j+1)$ with probability α, or to the site $(j-1)$ with probability $\beta = (1-\alpha)$. If $\alpha \neq \beta$, the random walk is biased. A bias models the effect of a constant force acting on the random walker, causing a systematic drift toward the left (if $\alpha < \beta$) or the right (if $\alpha > \beta$). The difference between a discrete-time

random walk in which the walker hops to a neighbouring site after a fixed interval of time (a 'time step'), and the continuous-time version under consideration here, is as follows: we now assume that the successive steps of the random walk are taken at *random* instants of time, with a mean jump rate λ. The random variable of current interest, however, is the *position* (or site label j) of the random walker at any time t. On an infinite linear lattice, we may take the random walker to start from the site $j = 0$ at $t = 0$, without loss of generality. What is the probability $P(j,t)$ that she is at the site j at time t? The crucial technical assumption is that the jump events themselves constitute a Poisson birth process—that is, the number of steps taken in any given time interval t is Poisson-distributed with a mean value λt. This makes $j(t)$ a Markov process. Moreover, since j can both increase as well as decrease with time as the walker moves back and forth, $j(t)$ is an example of a birth-and-death process.

As before, we write down an equation for $P(j, t+\delta t)$ based on the assumption that successive steps are independent of each other, and that δt is small enough to ensure that only three things can possibly happen in that incremental time interval: namely, the walker (i) either takes a step to the right, (ii) or takes a step to the left, (iii) or does not take a step at all. These three possibilities are mutually exclusive events. Hence their contributions to $P(j, t+\delta t)$ add up, and we get

$$P(j,t+\delta t) = (\lambda\,\alpha\,\delta t)\,P(j-1,t) + (\lambda\,\beta\,\delta t)\,P(j+1,t) + (1-\lambda\,\delta t)\,P(j,t).$$

In the limit $\delta t \to 0$, this yields the master equation

$$\frac{dP(j,t)}{dt} = \lambda\,\big\{\alpha\,P(j-1,t) - P(j,t) + \beta\,P(j+1,t)\big\}, \quad j \in \mathbb{Z}.$$

The initial condition is of course $P(j,0) = \delta_{j,0}$. Once again, it is convenient to solve the master equation using the generating function of $P(j,t)$.

★ **35.** Define the generating function $f(z,t) = \sum_{j=-\infty}^{\infty} P(j,t)\,z^j$. Note that the summation over j runs over *all* integers (and not just the non-negative ones).

(a) Show that $f(z,t)$ satisfies the first-order differential equation

$$\frac{\partial f}{\partial t} = \lambda\,(\alpha z - 1 + \beta z^{-1})\,f.$$

(b) Use the initial condition $f(z,0) = 1$ to obtain

$$f(z,t) = \exp\{\lambda t\,(\alpha z - 1 + \beta z^{-1})\}.$$

(c) You need the coefficient of z^j in the expansion of $f(z,t)$ in integer powers of z. Recall that the generating function of the modified Bessel function of the first kind, $I_j(u)$, is given by

$$\exp\left\{\tfrac{1}{2}u\left(z+z^{-1}\right)\right\} = \sum_{j=-\infty}^{\infty} I_j(u)\, z^j.$$

Hence read off the result

$$P(j,t) = e^{-\lambda t}\,(\alpha/\beta)^{j/2}\, I_j\big(2\lambda t\sqrt{\alpha\beta}\,\big).$$

This solution is valid for all integer values of j—positive, negative and zero. Since $I_j(u) \equiv I_{-j}(u)$ when j is an integer, the effect of the bias in the random walk is essentially carried by the factor $(\alpha/\beta)^{j/2}$ in the solution above. It is obvious that, when $\alpha > \beta$, the probability is larger for positive values of j than it is for negative values. The situation is reversed when $\beta > \alpha$, as expected.

★ **36. *Inclusion of a sojourn probability*:** Here's a slightly more generalized version of the biased random walk discussed above. Suppose the walker jumps from any site j to neighboring sites $j+1$ and $j-1$ with respective probabilities α and β, as before, and *stays* at the site j with a probability γ, where $\alpha+\beta+\gamma = 1$. What is $P(j,t)$ in this case?

Hint: A nonzero stay or *sojourn* probability implies that a 'jump' $j \to j$ is made with an average rate $\lambda\,\gamma$. Show that the master equation is now given by

$$\frac{dP(j,t)}{dt} = \lambda\left\{\alpha\, P(j-1,t) - (1-\gamma)P(j,t) + \beta\, P(j+1,t)\right\}, \quad j \in \mathbb{Z}.$$

Define the generating function for $P(j,t)$ and proceed as before.

***Connection with the Skellam distribution*:** The expression obtained above for $P(j,t)$ in a biased random walk should look familiar to you. It is just the Skellam distribution for the *difference* of two independent Poisson-distributed random variables, with respective mean values $\mu = \alpha\lambda t$ and $\nu = \beta\lambda t$, so that $\mu+\nu = \lambda t$. In other words, the position j of the random walker in a biased random walk on a linear lattice can be interpreted physically as the difference of two independent Poisson processes, namely, jumps to the right and left, respectively. The respective mean rates of these processes are $\alpha\lambda$ and $\beta\lambda$, so that the mean values of the individual processes in a time interval t are $\alpha\lambda t$ and $\beta\lambda t$. It is the Markov property of the random walk that leads to such a simple interpretation of the process.

From the known values of the mean and variance of the Skellam distribution, we can write down the mean and variance of the displacement of the random walker at any time t. These are given by

$$\langle j(t)\rangle = (\alpha - \beta)\lambda t, \quad \text{Var}\, j(t) \stackrel{\text{def.}}{=} \langle j^2(t)\rangle - \langle j(t)\rangle^2 = \lambda t.$$

The first of the equations above shows how a bias in the random walk leads to a drift, as measured by the mean displacement, that increases linearly with time. The second shows how the expected *diffusive behavior* of the random walk emerges, once the effect of the drift is subtracted out from the mean squared displacement. This is the most important feature of the random walk.

***Asymptotic behavior of the probability*:** The fact that the variance of $j(t)$ becomes unbounded as $t \to \infty$ suggests that the probability $P(j,t)$ does not, in fact, tend to any stationary distribution in that limit. In order to see this explicitly, consider the asymptotic ($t \to \infty$) behavior of the exact solution for $P(j,t)$. What you need for this purpose is the leading asymptotic behavior of the modified Bessel function when its argument tends to infinity. This is given by

$$I_j(z) \sim \frac{e^z}{\sqrt{2\pi z}} \quad \text{as} \quad |z| \to \infty,$$

for all finite values of j. It follows that the leading asymptotic behavior of $P(j,t)$ as $t \to \infty$ is given by

$$P(j,t) \sim \frac{(\alpha/\beta)^{j/2}\, e^{-\lambda t(1-2\sqrt{\alpha\beta})}}{\left(4\pi\lambda t\sqrt{\alpha\beta}\,\right)^{1/2}}\,.$$

But $1 - 2\sqrt{\alpha\beta} = 1 - 2\sqrt{\alpha(1-\alpha)} \geq 0$. The equality sign is attained only when $\alpha = \beta = \frac{1}{2}$. We conclude that:

(a) $P(j,t)$ decays *exponentially* to zero as $t \to \infty$ for all $\alpha \neq \frac{1}{2}$, i.e., when the random walk is biased (either to the right or to the left).

(b) When the random walk is unbiased ($\alpha = \beta = \frac{1}{2}$), this asymptotic behavior is drastically modified. $P(j,t)$ now decays to zero as $t \to \infty$ like an inverse *power* of t, namely, like $1/\sqrt{t}$.

(c) In either case, $P(j,t)$ vanishes in the limit $t \to \infty$, for all j. Hence the random walk process has *no* stationary distribution.

3.11 Sedimentation under gravity: the barometric distribution

A direct application of a biased random walk is provided by a particle inn a vertical column of fluid, subject to the (constant) force of gravity. Modeling

the random motion of the particle by a biased random walk over the sites $j = 0, 1, 2, \ldots$ stacked vertically from 0 upwards, we have here a case when $\beta > \alpha$ owing to the downward drift under gravity. The stationary distribution $P(j)$ is obtained by setting $dP(j,t)/dt = 0$ in the master equation. We find

$$\alpha\, P(j-1) - P(j) + \beta\, P(j+1) = 0, \quad \text{for } \; j \geq 1.$$

For $j = 0$, however, we have a different equation. Since there is no loss of probability at the 'floor' $j = 0$ (which acts like a **reflecting boundary**), we have

$$\alpha\, P(0) - \beta\, P(1) = 0.$$

★ 37. The foregoing set of equations can be solved for the stationary distribution $P(j),\ j \geq 0$. Show that the normalised distribution is

$$P(j) = \frac{(\beta - \alpha)}{\beta} \left(\frac{\alpha}{\beta}\right)^j.$$

This is a geometric distribution.

The probability $P(j)$ decreases exponentially with increasing j, since it can be written as

$$P(j) = \frac{(\beta - \alpha)}{\beta}\, e^{-j \,\ln(\beta/\alpha)}.$$

Later on, we will pass to the continuum limit of the one-dimensional diffusion of a particle under a constant force, which describes the process of sedimentation. We'll see that the expression for $P(j)$ found here reduces precisely to the barometric distribution of the density of the atmosphere, under the simplifying assumption that the temperature is uniform.

3.12 General birth-and-death process

A general birth-and-death process involves transition probabilities (or rates) that are *state-dependent*. The master equation in this case is a generalization of that for a biased random walk. Absorbing the rate parameter λ into the bias probabilities α and β, we now have

$$\frac{dP(j,t)}{dt} = \alpha_{j-1}\, P(j-1,t) - (\alpha_j + \beta_j)\, P(j,t) + \beta_{j+1}\, P(j+1,t).$$

The functions α_j and β_j now have the physical dimensions of $[\text{time}]^{-1}$. An analytic closed-form solution for $P(j,t)$ is no longer possible, in general. In the important case when α_j and β_j are *linear* functions of j, however, such a solution may be found.

The master equation above is useful in modelling a large number of physical phenomena. One is interested, in particular, in the stationary distribution, if any, to which $P(j,t)$ tends in the limit $t \to \infty$. This stationary distribution satisfies the difference equation

$$\alpha_{j-1}\, P(j-1) - (\alpha_j + \beta_j)\, P(j) + \beta_{j+1}\, P(j+1) = 0.$$

It is clear that the integer-valued random variable j can have essentially three different possible ranges, that is (i) infinite in both directions, or (ii) infinite in one direction, or (iii) bounded from both above and below. In cases (ii) and (iii), there is the possibility of a stationary distribution, depending on the boundary condition imposed at the finite end(s). Consider, for definiteness, the case when $j \geq 0$. The master equation for $P(0,t)$ is different from that for $P(j,t),\ j \geq 1$. If there is no 'leakage of probability' at $j = 0$, then we must have

$$\frac{dP(0,t)}{dt} = -\alpha_0\, P(0,t) + \beta_1\, P(1,t).$$

Correspondingly, the equation for the stationary distribution becomes

$$-\alpha_0\, P(0) + \beta_1\, P(1) = 0.$$

★ **38.** Use the foregoing to solve successively for $P(1), P(2), \ldots$ in terms of $P(0)$, to obtain

$$P(j) = \frac{\alpha_0\, \alpha_1\, \cdots\, \alpha_{j-1}}{\beta_1\, \beta_2\, \cdots\, \beta_j}\, P(0).$$

Hence, assuming that the infinite series in the denominator converges, show that

$$P(j) = \Big\{\frac{\alpha_0\, \alpha_1\, \cdots\, \alpha_{j-1}}{\beta_1\, \beta_2\, \cdots\, \beta_j}\Big\} \Big/ \Big\{1 + \sum_{k=1}^{\infty} \frac{\alpha_0\, \alpha_1\, \cdots\, \alpha_{k-1}}{\beta_1\, \beta_2\, \cdots\, \beta_k}\Big\}.$$

When the upper limit on j is some finite integer N, of course, the sum in the denominator runs up to $k = N$.

A couple of physical applications of this result are as follows:

***Blackbody radiation*:** 'Quantization' of the electromagnetic field involves a mathematical equivalence between the field and a set of linear harmonic oscillators, one for each wave vector and state of polarization. An oscillator with an energy level spacing equal to $h\nu$ can interact with radiation of frequency ν by absorbing and emitting photons. Absorption (respectively, emission) of a photon takes the oscillator from the level j to the level $j+1$ (respectively, $j-1$). It turns out that $\alpha_{j-1} = Aj$ and $\beta_j = Bj$ in this case, where A and B are positive constants. It follows that $P(j) \propto (A/B)^j$. Identifying A/B with $e^{h\nu/k_B T}$ then

leads to Planck's formula for the frequency distribution of blackbody radiation.

***A chemical reaction*:** Consider the chemical reaction $A \rightleftharpoons B$ with rate constants K and K' for the forward and backward reactions. Let the number of molecules of A be so large that it essentially remains unchanged by the on-going reaction. Then, if j is the number of molecules of the product B, we have $\alpha_j \approx \alpha$, a constant, while $\beta_j = K'j$. It follows that $P(j)$ is a Poisson distribution with mean value α/K'.

4 Continuous Markov processes

4.1 Master equation for the conditional density

We turn now to the case of a continuous Markov process $X(t)$. The sample space of the random variable is some continuous set of values. The preceding discussion of discrete Markov processes can be extended to the continuous case with obvious modifications, such as the replacement of probabilities by the corresponding probability density functions. Thus, the n-time joint probability density for a stationary, continuous Markov process is given by

$$p_n(x_n\,,\,t_n\,;\,x_{n-1}\,,\,t_{n-1}\,;\,\ldots\,;\,x_1\,,\,t_1) = \Big\{\prod_{r=1}^{n-1} p(x_{r+1}\,,\,t_{r+1}-t_r\,|\,x_r)\Big\}\,p(x_1),$$

for every $n \geq 2$. (Here, the values $\{x_k\}$ belong to the sample space of $X(t)$.) Hence the fundamental quantity characterizing a stationary continuous Markov process is the conditional density $p(x, t\,|\,x_0)$. The stationary PDF $p(x)$ is expected to be related to this quantity according to

$$\lim_{t\to\infty}\,p(x\,,\,t\,|\,x_0) = p(x),$$

independent of the initial value x_0 . As in the discrete case, the conditional density satisfies the Chapman-Kolmogorov equation

$$p(x\,,\,t\,|\,x_0) = \int dx'\,p(x\,,\,t-t'\,|\,x')\;p(x'\,,\,t'\,|\,x_0) \quad (0 \leq t' \leq t).$$

The integration runs over the range of values assumed by the random variable.

As before, this nonlinear equation may be converted to a linear one. Let $w(x\,|\,x')\,dx$ be the probability per unit time of a transition from a given value x' of the random variable to any value in the range $(x\,,\,x+dx)$. Therefore $w(x\,|\,x')$ is the **transition probability density** per unit time. The chain equation can then be reduced to the master equation

$$\frac{\partial}{\partial t}\,p(x\,,\,t\,|\,x_0) = \int dx'\,\Big\{p(x',\,t\,|\,x_0)\,w(x\,|\,x') - p(x\,,\,t\,|\,x_0)\,w(x'\,|\,x)\Big\}.$$

The initial condition is obviously $p(x\,,\,0\,|\,x_0) = \delta(x-x_0)$.

★ **39.** Using the same procedure as in the discrete case, derive the master equation above from the Chapman-Kolmogorov equation.

4.2 The Fokker-Planck equation

Even though the master equation is a linear equation for the conditional PDF $p(x\,,\,t\,|\,x_0)$, it is an *integro-differential* equation. This is the price we pay for the reduction of the *nonlinear* Chapman-Kolmogorov to a *linear* equation. Solving it exactly is far from an easy task. One approach is to convert it to a partial differential equation.

★ **40.** Show that the master equation can be written in the form

$$\frac{\partial}{\partial t}\,p(x\,,\,t\,|\,x_0) = \sum_{n=0}^{\infty}(-1)^n\frac{\partial^n}{\partial x^n}\,[A_n(x)\,p(x\,,\,t\,|\,x_0)],$$

where

$$A_n(x) = \int dx'\,x'^{\,n}\,w(x+x'\,|\,x).$$

Hint: Write $w(x\,|\,x') = w(x'+\delta x\,|\,x')$, and expand this function as a Taylor series in power of δx. Integration by parts then leads to the result quoted.

This is called the **Kramers-Moyal expansion**. While it is formally equivalent to the master equation, the problem is that, in general, it is of *infinite* order in the partial derivative $\partial/\partial x$. In many physical applications, however, the Kramers-Moyal form of the master equation either reduces to, or can be well-approximated by, the *second-order* partial differential equation

$$\frac{\partial}{\partial t}\,p(x\,,\,t\,|\,x_0) = -\frac{\partial}{\partial x}\,[A_1(x)\,p(x\,,\,t\,|\,x_0)] + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\,[A_2(x)\,p(x\,,\,t\,|\,x_0)].$$

This is known as the **forward Kolmogorov equation** or, more commonly in the physics literature, the **Fokker-Planck equation**. There is an interesting *exact* result in this regard: the Kramers-Moyal expansion either terminates at the second order, or else is an infinite series—nothing in between. Of course truncation at the second order may be a good *approximation* in the latter case, depending on the problem at hand. The Fokker-Planck equation is the most frequently used form of the master equation for the conditional PDF in physical applications of continuous Markov processes. In the mathematical literature on stochastic processes, all continuous Markov processes whose conditional densities satisfy the Fokker-Planck equation are called **diffusion processes**. The functions

$$A_1(x) = \int dx'\,x'\,w(x+x'\,|\,x), \quad A_2(x) = \int dx'\,x'^{\,2}\,w(x+x'\,|\,x),$$

which are essentially the first two moments of the transition rate, are are referred to as the **drift coefficient** and **diffusion coefficient**, respectively. The solution of the Fokker-Planck equation itself is again a nontrivial task. Part of the

technical problem is that the differential operator on the right-hand side is not a self-adjoint operator. There exists a considerable body of literature on the analysis of the Fokker-Planck equation and its solution in various cases. The familiar diffusion equation, as well several other related equations to be discussed below, are examples of Fokker-Planck equations.

Some features of the equation, however, can be deduced quite easily. The Fokker-Planck equation can be written in the form of an *equation of continuity*:

$$\frac{\partial}{\partial t}\, p(x\,,\, t\,|\, x_0) + \frac{\partial}{\partial x}\, j(x\,,\, t\,|\, x_0) = 0,$$

where the probability current density j is given by

$$j(x\,,\, t\,|\, x_0) = -\frac{\partial}{\partial x}\left\{\tfrac{1}{2} A_2(x)\, p(x\,,\, t\,|\, x_0)\right\} + A_1(x)\, p(x\,,\, t\,|\, x_0).$$

Consider, now, what happens as $t \to \infty$. We have $p(x\,,\, t\,|\, x_0) \to p(x)$ (assuming that a stationary PDF $p(x)$ exists). Correspondingly, $j(x,t)$ tends to the stationary current

$$j^{\rm st}(x) = -\frac{d}{dx}\left\{\tfrac{1}{2} A_2(x)\, p(x)\right\} + A_1(x)\, p(x).$$

Obviously, $\partial p(x)/\partial t = 0$. Hence $dj^{\rm st}/dx = 0$, so that $j^{\rm st}(x)$ is actually a constant (independent of x). This means that the stationary density itself can be found by solving the *first-order, ordinary* differential equation

$$\frac{d}{dx}\left\{\tfrac{1}{2} A_2(x)\, p(x)\right\} - A_1(x)\, p(x) = \text{ constant},$$

where the constant is obtained from the boundary conditions in any given instance.

4.3 The autocorrelation function for a continuous process

Before we turn to a very important continuous, stationary Markov process described by the Fokker-Planck equation, and specific physical examples thereof, it's useful to write down the general expression for the autocorrelation function in the case of a stationary, continuous Markov process. The formulas that follow are straightforward extensions of those already written down for a stationary discrete stochastic process. We have

$$\begin{aligned}
\langle X(t')\, X(t)\rangle &= \int dx' \int dx\; x\, x'\; p_2(x\,, t\,;\, x'\,,\, t') \\
&= \int dx' \int dx\; x\, x'\; p_2(x\,, t\,|\, x'\,,\, t')\, p_1(x'\,,\, t').
\end{aligned}$$

For a stationary process this becomes, on setting $t' = 0$,

$$\langle X(0)\, X(t)\rangle = \int dx' \int dx\ x\, x'\ p(x\,, t\,|\,x')\, p(x'),$$

in terms of the stationary and conditional PDFs. This is the autocorrelation function $C_X(t)$ when the mean value of $X(t)$ is zero.

When the mean value of the random process is nonzero, the autocorrelation function is defined in terms of the deviation $\delta X = X - \langle X\rangle$ from the mean value. We then have

$$C_X(t) \stackrel{\text{def.}}{=} \langle \delta X(0)\, \delta X(t)\rangle = \int dx' \int dx\ x\, x'\ p(x\,, t\,|\,x')\, p(x') - \langle X\rangle^2,$$

where

$$\langle X\rangle = \int dx\, x\, p(x).$$

4.4 The Ornstein-Uhlenbeck process

Among stationary, continuous Markov processes, there is a very unique one:

- There is only one continuous random process that is stationary, Markov, as well as Gaussian. This is the **Ornstein-Uhlenbeck process**.

By a 'Gaussian process' we mean that all the joint probability densities of the random variable are Gaussians in functional form. It is understood that the range of the random variable is $(-\infty, \infty)$.

The Ornstein-Uhlenbeck (OU, for short) process corresponds to the case when the drift coefficient is proportional to x, and the diffusion coefficient is a constant:

$$A_1(x) = -a_1\, x \ \text{ and } \ A_2(x) = a_2\,, \ \text{ where } \ a_1, a_2 = \text{positive constants.}$$

We'll discuss two physical examples of this random process, shortly. The Fokker-Planck equation in this case is

$$\frac{\partial}{\partial t}\, p(x\,,\, t\,|\,x_0) = a_1 \frac{\partial}{\partial x}\, [x\, p(x\,,\, t\,|\,x_0)] + \frac{a_2}{2}\, \frac{\partial^2}{\partial x^2}\, p(x\,,\, t\,|\,x_0).$$

The physical dimensions of a_1 and a_2 are, respectively, 1/[time] and $[x^2]$/[time]. The initial condition on $p(x\,,\, t\,|\,x_0)$ is of course given by

$$p(x\,,\, 0\,|\,x_0) = \delta(x - x_0).$$

The simplest case corresponds to the natural boundary conditions

$$p(x\,,\, t\,|\,x_0) = 0 \quad \text{as} \quad x \to \pm\infty.$$

The solution that satisfies these conditions will be written down subsequently. The main physical features of this solution are as follows:

(i) The conditional density starts at $t = 0$ as a δ-function peak at x_0, and widens as t increases. It is a Gaussian for all $t > 0$, and attains a limiting Gaussian form as $t \to \infty$.

(ii) As t increases, the mean value of $X(t)$ (and hence the peak of the Gaussian) drifts monotonically to zero like a decaying exponential function of t.

(iii) Simultaneously, the variance of $X(t)$ increases monotonically from zero, and tends to a saturation value as $t \to \infty$.

(iv) The autocorrelation function of $X(t)$ decays to zero exponentially in time, with a single characteristic time constant given by $1/a_1$.

★ **41.** Some of these features can be deduced even without solving the Fokker-Planck equation explicitly. Consider, for instance, the mean value $\langle X(t)\rangle$ and the second moment $\langle X^2(t)\rangle$. These are defined as

$$\mu_X(t) \equiv \langle X(t)\rangle \stackrel{\text{def.}}{=} \int_{-\infty}^{\infty} dx\, x\; p(x\,,\, t\,|\,x_0) \;\text{ and }\; \langle X^2(t)\rangle \stackrel{\text{def.}}{=} \int_{-\infty}^{\infty} dx\, x^2\; p(x\,,\, t\,|\,x_0),$$

respectively.

(a) These moments satisfy ordinary first-order differential equations in time. Show that

$$\frac{d}{dt}\langle X(t)\rangle + a_1\langle X(t)\rangle = 0 \quad \text{and} \quad \frac{d}{dt}\langle X^2(t)\rangle + 2a_1\langle X^2(t)\rangle - a_2 = 0.$$

(b) The moments under discussion are *conditional* averages: that is, they are averages over all realizations of the random process $X(t)$, *given* the specific initial value x_0. Therefore the initial conditions on the moments are simply

$$\langle X(0)\rangle = x_0 \quad \text{and} \quad \langle X^2(0)\rangle = x_0^2\,.$$

Show that the solutions for the first and second moments are given by

$$\langle X(t)\rangle \equiv \mu_X(t) = x_0\, e^{-a_1 t} \quad \text{and} \quad \langle X^2(t)\rangle = \frac{a_2}{2a_1} + \Big(x_0^2 - \frac{a_2}{2a_1}\Big)e^{-2a_1 t},$$

for all $t \geq 0$.

Hint: Multiply both sides of the Fokker-Planck equation by x and x^2, respectively, and integrate over x. Note that

$$\frac{d\langle X(t)\rangle}{dt} = \int_{-\infty}^{\infty} dx\, x\, \frac{\partial p}{\partial t} \;\text{ and }\; \frac{d\langle X^2(t)\rangle}{dt} = \int_{-\infty}^{\infty} dx\, x^2\, \frac{\partial p}{\partial t}\,.$$

Use the Fokker-Planck equation for $\partial p/\partial t$, and integrate by parts. All the boundary terms at $x = \pm\infty$ may be set equal to zero, since $p(x\,,\,t\,|\,x_0)$ and its derivatives with respect to x tend to zero faster than any inverse power of x as $|x| \to \infty$. (This will be explicit in the solution to be written down shortly.)

The time-dependent variance of $X(t)$ is

$$\mathrm{Var}\,X(t) \equiv \sigma_X^2(t) = \langle X^2(t)\rangle - \mu_X^2(t) = \frac{a_2}{2a_1}\big(1 - e^{-2a_1 t}\big).$$

The variance thus increases monotonically from its initial value 0 to the limiting value $a_2/(2a_1)$. Observe that it is actually independent of the initial value x_0 for all t.

4.5 The Ornstein-Uhlenbeck distribution

Let's now consider the exact solution of the Fokker-Planck equation

$$\frac{\partial}{\partial t}\,p(x\,,\,t\,|\,x_0) = a_1\frac{\partial}{\partial x}\,[x\,p(x\,,\,t\,|\,x_0)] + \frac{a_2}{2}\,\frac{\partial^2}{\partial x^2}\,p(x\,,\,t\,|\,x_0).$$

with the initial condition $p(x\,,\,0\,|\,x_0) = \delta(x - x_0)$ and natural boundary conditions $p(x\,,\,t\,|\,x_0) = 0$ as $x \to \pm\infty$. There are several ways of solving the equation. I will not discuss these here, but will merely write down the solution, based on the assertion that the PDF is a Gaussian at all times. A knowledge of the mean and variance therefore suffices for our purposes, because a Gaussian is fully determined by its mean and variance. The normalized conditional probability density is given by

$$p(x\,,\,t\,|\,x_0) = \frac{1}{\sqrt{2\pi\sigma_X^2(t)}}\,\exp\,\Big\{-\frac{\big(x - \mu_X(t)\big)^2}{2\sigma_X^2(t)}\Big\},$$

where, as already found, the mean and variance are given by

$$\mu_X(t) = x_0\,e^{-a_1 t} \quad \text{and} \quad \sigma_X^2(t) = \frac{a_2}{2a_1}\big(1 - e^{-2a_1 t}\big).$$

This is the **Ornstein-Uhlenbeck distribution** (or rather, the corresponding PDF). The claim is that it is the PDF of the only continuous stochastic process that is stationary, Markov, as well as Gaussian. As $t \to \infty$, this conditional PDF tends to the stationary PDF

$$p(x) = \sqrt{a_1/(\pi a_2)}\;e^{-a_1 x^2/a_2}.$$

The autocorrelation function of the Ornstein-Uhlenbeck process has already been stated to be a decaying exponential function of t. Let's see how this comes about.

The first step is to recognize that the mean value of the process, as opposed to the *conditional* mean $\mu_X(t) = \langle X(t) \rangle = x_0\, e^{-a_1 t}$ found above, is actually zero. This is easily done, because all that is needed is a *further* averaging of $\langle X(t) \rangle$ over all initial conditions weighted with the stationary density $p(x_0)$. But the latter is a Gaussian centered at $x_0 = 0$, and hence a symmetric function of x_0. Therefore the full, or unconditional, average of X is zero. The autocorrelation function is then given by

$$C_X(t) = \langle X(0)\, X(t) \rangle = \int_{-\infty}^{\infty} dx_0 \int_{-\infty}^{\infty} dx\ x\, x_0\ p(x\,, t \,|\, x_0)\, p(x_0).$$

★ **42.** Use the expressions found above for $p(x\,,\, t \,|\, x_0)$ and $p(x_0)$ and carry out the integrations over x and x_0 to obtain

$$C_X(t) = \frac{a_2}{2a_1}\, e^{-a_1 t}.$$

Hint: The calculation becomes much easier if you carry out the integration over x first. Shift the variable of integration from x to $x - \mu_X(t)$.

Doob's Theorem: This result establishes the assertion made earlier, namely, that the autocorrelation function of the OU process is a decaying exponential function of time, with a correlation time given by $1/a_1$. There is an interesting exact result in this regard, known as **Doob's Theorem**:

- The only continuous, stationary, Gaussian process with an exponentially decaying autocorrelation is the OU process.

All the qualifying clauses are necessary, as omitting any one of them yields counter-examples. For instance:

—The dichotomous Markov process is stationary and exponentially correlated, but it is a jump process and not a continuous process. It is certainly not Gaussian.

—The Wiener process (or Brownian motion), whose probability density satisfies the standard diffusion equation, is continuous, Gaussian and Markov, but not stationary. It is not exponentially correlated.

—There is a whole class of stochastic processes that are continuous, stationary, Markov and exponentially-correlated. This class of processes includes, but is not restricted to, Gaussian Markov processes. The Gaussian member of this class of processes is precisely the OU process. These processes have stationary PDFs that satisfy the differential equation

$$\frac{dp(x)}{dx} = \frac{A(x)}{B(x)}\, p(x),$$

where $A(x)$ and $B(x)$ are polynomials in x of degree no greater than 1 and 2, respectively. Further, (i) the solution $p(x)$ must have finite first and second moments, and (ii) $B(x)p(x)$ must vanish at the end-points of the range of x.

★ **43.** Setting $A(x) = a_0 + a_1\, x$ and $B(x) = b_0 + b_1\, x + b_2\, x^2$, show that five distinct nontrivial possibilities arise. (Clearly, the Gaussian corresponds to the case $a_0 = 0, a_1 < 0, b_0 > 0, b_1 = b_2 = 0$.)

4.6 Velocity distribution in a gas

The Fokker-Planck equation arose originally in the context of the velocity distribution of the molecules of a classical ideal gas in thermal equilibrium at a temperature T. From elementary statistical physics, you know that the normalized stationary PDF of each Cartesian component of the velocity of a molecule is a Gaussian. For notational simplicity, let us denote any one of the Cartesian components of the velocity of a molecule by U (or u, depending on whether we are talking about the random variable or its value). The stationary or equilibrium PDF of U is given by

$$p(u) = \Big(\frac{m}{2\pi k_B T}\Big)^{1/2} \exp\Big\{-\frac{mu^2}{2k_B T}\Big\},$$

where m is the mass of a molecule and k_B is Boltzmann's constant. It is quite natural to ask a related question: Suppose we focus on any one particular molecule (the 'tagged' particle), and find that its instantaneous velocity component at $t = 0$ is equal to u_0. How does its velocity distribution change with time, and attain the Maxwellian form as $t \to \infty$? This simple-looking question already takes us beyond the purview of *equilibrium* statistical mechanics, as it involves the *time-dependent* conditional probability density $p(u\,,\, t \,|\, u_0)$.

The simplest model that describes the physical situation is based on a random or *stochastic differential equation* for the velocity of the tagged particle, called the **Langevin equation**. It then turns out that the velocity component $U(t)$

of the tagged particle is a stationary, Gaussian, Markov process. Its conditional probability density satisfies the Fokker-Planck equation

$$\frac{\partial}{\partial t}\, p(u\,,\, t\,|\, u_0) = \gamma\, \frac{\partial}{\partial u}\, [u\, p(u\,,\, t\,|\, u_0)] + \frac{\gamma k_B T}{m}\, \frac{\partial^2}{\partial u^2}\, p(u\,,\, t\,|\, u_0).$$

Here γ is a positive constant with the physical dimensions of $[\text{time}]^{-1}$. It is directly proportional to the viscosity of the fluid. It is evident that $U(t)$ is an Ornstein-Uhlenbeck process, with

$$a_1 = \gamma \quad \text{and} \quad a_2 = 2\gamma k_B T/m\,.$$

The conditional mean and variance of the velocity are therefore given by

$$\mu_U(t) = u_0\, e^{-\gamma t} \quad \text{and} \quad \sigma_U^2(t) = \frac{k_B T}{m}\, \big(1 - e^{-2\gamma t}\big).$$

We can now write down the normalized fundamental solution to the Fokker-Planck equation with the initial condition

$$p(u\,,\, 0\,|\, u_0) = \delta(u - u_0)$$

and natural boundary conditions

$$p(u\,,\, t\,|\, u_0) \to 0 \quad \text{as} \quad u \to \pm\infty.$$

It is the Ornstein-Uhlenbeck distribution

$$p(u\,,\, t\,|\, u_0) = \Big\{\frac{m}{2\pi k_B T\big(1 - e^{-2\gamma t}\big)}\Big\}^{1/2} \exp\, \Big\{-\, \frac{m\big(u - u_0\, e^{-\gamma t}\big)^2}{2k_B T\big(1 - e^{-2\gamma t}\big)}\Big\}.$$

★ **44.** Verify that the expression above satisfies the Fokker-Planck equation.

The autocorrelation function of the velocity component is, as expected, a decaying exponential in t. Setting $a_1 = \gamma$ and $a_2 = 2\gamma k_B T/m$ in the expression for the autocorrelation function of the Ornstein-Uhlenbeck process, we get

$$C_U(t) = \frac{k_B T}{m}\, e^{-\gamma t} \quad (t \geq 0).$$

It is immediately evident that $1/\gamma$ is the correlation time of the velocity process. It is the characteristic relaxation time (or **equilibration time**) over which the velocity 'thermalizes', starting from any specified initial value u_0.

Finally, in the interests of technical accuracy I must mention that the foregoing model (as it stands) is too simplistic to be directly applicable to an individual molecule as the tagged particle. The Langevin equation and the associated Fokker-Planck equation are actually more appropriate for describing the random motion of a much more massive tagged particle moving in a fluid of much less massive molecules.

4.7 Solution for an arbitrary initial velocity distribution

The solution of the Fokker-Planck equation written down above corresponds to a 'sharp' initial condition, namely, $p(u\,,\,0\,|\,u_0) = \delta(u - u_0)$. This means that it is, in fact, the fundamental **Green function** for the differential operator in the Fokker-Planck equation. Hence it can be used to write down the solution for an *arbitrary* initial distribution of velocities, because the Fokker-Planck equation is a *linear* equation.

Let the initial distribution be given by the normalized PDF $p_{\rm init}(u_0)$. Then the PDF of the velocity component at any time $t > 0$ is given by the expression

$$p(u,t) = \Big\{\frac{m}{2\pi k_B T\big(1 - e^{-2\gamma t}\big)}\Big\}^{1/2} \int_{-\infty}^{\infty} du_0 \, \exp\Big\{-\,\frac{m\big(u - u_0\, e^{-\gamma t}\big)^2}{2k_B T\big(1 - e^{-2\gamma t}\big)}\Big\} p_{\rm init}(u_0).$$

This solution exhibits two interesting properties:

(i) Regardless of the initial density $p_{\rm init}(u_0)$, the solution tends to the stationary Gaussian PDF $p(u)$ as $t \to \infty$.

(ii) Suppose the initial probability density is the Maxwellian distribution itself, i.e., $p_{\rm init}(u_0) = p(u_0)$. Then $p(u,t)$ remains equal to $p(u)$ at *all* times.

★ **45.** Verify the statements (i) and (ii) above. They show how robust and stable the state of thermal equilibrium is.

4.8 Diffusion of a harmonically bound particle

Another physical example of the Ornstein-Uhlenbeck distribution arises from the diffusion equation for a harmonically bound particle. The particle undergoes Brownian motion on a line, the x-axis, say, while it is under the influence of a harmonic oscillator potential $\frac{1}{2}m\omega^2 x^2$. (The natural frequency of the oscillator is ω.) The particle is also subject to a frictional force $-m\gamma\dot{x}(t)$ arising from the medium in which it moves (a fluid in thermal equilibrium at temperature T), where γ is a positive constant with the physical dimensions of $[\text{time}]^{-1}$. We are interested in the conditional PDF $p(x,t\,|\,x_0)$ of the position $X(t)$ of the particle, given that it starts at $t = 0$ from some point x_0.

Recall from elementary physics that a linear harmonic oscillator is underdamped when $\gamma < 2\omega$, and overdamped when $\gamma > 2\omega$. It turns out that, in the *over*damped case, and at sufficiently long times (specifically, for $\gamma t \gg 1$), the PDF $p(x,t\,|\,x_0)$ satisfies the following partial differential equation:

$$\frac{\partial}{\partial t}\, p(x,t\,|\,x_0) = \frac{\omega^2}{\gamma}\, \frac{\partial}{\partial x}\big[x\, p(x,t\,|\,x_0)\big] + \frac{k_B T}{m\gamma}\, \frac{\partial^2}{\partial x^2}\, p(x,t\,|\,x_0).$$

This is an example of a diffusion equation for the positional PDF in the presence of an applied force or in an external potential. (Such an equation is called a Smoluchowski equation.) The partial differential equation satisfied by $p(x, t \,|\, x_0)$ is clearly a Fokker-Planck equation, with coefficients given by

$$a_1 = \omega^2/\gamma \quad \text{and} \quad a_2 = 2k_B T/(m\gamma)\,.$$

Thus $X(t)$ is an Ornstein-Uhlenbeck process, under the conditions mentioned above. For this reason, the Ornstein-Uhlenbeck process itself is sometimes called the **oscillator process**.

The fundamental solution of the Fokker-Planck equation for $p(x, t \,|\, x_0)$ is a Gaussian with conditional mean and variance given by

$$\mu_X(t) = x_0\, e^{-\omega^2 t/\gamma} \quad \text{and} \quad \sigma_X^2(t) = \frac{k_B T}{m\omega^2}\left(1 - e^{-2\omega^2 t/\gamma}\right).$$

Letting $t \to \infty$ in this solution, we get

$$\lim_{t\to\infty} p\,(x, t\,|x_0) = p(x) = \left(\frac{m\omega^2}{2\pi k_B\, T}\right)^{1/2} \exp\left\{-\,\frac{m\omega^2 x^2}{2k_B T}\right\}.$$

This is just the normalized stationary or equilibrium PDF that we would write down in the canonical ensemble in equilibrium statistical mechanics.

I reiterate that the Fokker-Planck equaltion and its Gaussian solution written down above are not exact relations for the positional PDF $p(x, t \,|\, x_0)$ of a harmonically bound diffusing particle. They are approximations that are only valid in the highly overdamped case, and at times $t \gg \gamma^{-1}$. Only under these circumstances does the position variable $X(t)$ become a stationary Markov process, with an autocorrelation function given by the decaying exponential

$$C_X(t) = \frac{k_B T}{m\omega^2}\, e^{-\omega^2 t/\gamma}.$$

Note, too, that the existence of the stationary Gaussian PDF $p(x)$ has another implication. It means that, in stark contrast to a free particle, a harmonically bound particle in thermal equilibrium with a heat bath does not undergo any long-range diffusion at all! The variance of its displacement does not increase linearly with t — in fact, it does not diverge like any power of t. Instead, it saturates (as $t \to \infty$) to the value $k_B T/m\omega^2$. This expression is precisely what you would write down based on an elementary application of the equipartition theorem, according to which $\frac{1}{2}m\omega^2\langle x^2\rangle = \frac{1}{2}k_B T$.

The quiz that follows covers both the topics discussed in the foregoing, as well as the topics dealt with in the rest of this course.

Quiz

I. Are the statements in quotation marks true or false?

1. Consider the set of n independent random variables $X_1, ..., X_n$, where X_j can take the value 1 with probability p_j, and the value -1 with probability $q_j = 1 - p_j$, with $0 < p_j < 1$. Let $Z_n = X_1 + X_2 + \cdots + X_n$.

   "The variance of Z_n is equal to $4 \sum_{j=1}^{n} p_j \, q_j$."

2. A random variable n takes values in the set of non-negative integers, and has a geometric distribution with mean value μ.

   "Every cumulant of n is equal to μ."

3. "If m and n are independent Poisson-distributed random variables, so is $\alpha \, m + (1 - \alpha) \, n$, where $0 < \alpha < 1$ is a constant."

4. "If μ and ν are the respective mean values of two independent Poisson-distributed random variables m and n, the variance of the random variable $m - n$ is $|\mu - \nu|$."

5. "If μ and ν are the respective mean values of two independent Poisson-distributed random variables m and n, the mean value of $(m + n)^2$ is equal to $(\mu + \nu)^2$."

6. Let X be a random variable, and let $Y = X + c$ where c is a constant.

   "Except for the first cumulant, every cumulant of Y is equal to the corresponding cumulant of X."

7. Let ξ be a random variable with a Gaussian distribution.

   "For every positive integer n, ξ can be written as the sum of n random variables, each with a Gaussian distribution."

8. Let X denote a random variable.

   "The quantity $\langle X^4 \rangle - 3\langle X^2 \rangle^2$ can never be negative."

9. Let X be a Gaussian random variable with zero mean and unit variance.

   "The normalized PDF $\rho(u)$ of the random variable $U = 1/X^2$ has a leading asymptotic behaviour proportional to $1/u^{3/2}$ as $u \to \infty$."

10. Let X_1, $X_2, \ldots$ denote a set of random variables.

    "If the correlation $\langle X_i \, X_j \rangle$ is proportional to δ_{ij}, then X_1, $X_2, \ldots$ must necessarily be statistically independent random variables."

11. The characteristic function of a random variable X with PDF $p(x)$ is defined as the Fourier transform $\widetilde{p}(k) = \int_{-\infty}^{\infty} dx \, e^{-ikx} \, p(x) \equiv \langle e^{-ikX} \rangle$.

    "If the imaginary part of $\widetilde{p}(k)$ vanishes identically, we may conclude that X and $-X$ are identically distributed."

12. "If $\widetilde{p}(k)$ is a characteristic function, so is the function $|\widetilde{p}(k)|^2$."

13. Let $X \in [0, \infty)$ be a random variable with a normalized PDF $p(x) = e^{-x}$.

    "The random variable $\xi = e^{-X}$ is uniformly distributed in $[0, 1]$."

14. A random variable X is Cauchy-distributed, with a PDF given by $p(x) = \lambda/[\pi(x^2 + \lambda^2)]$, where λ is a positive constant.

    "The reciprocal of X is also Cauchy-distributed."

15. "If a random variable can be written as the sum of n *iidrv* where n is a given positive integer, then it is k-divisible for all integer values of k in the range $1 \leq k \leq n$."

16. "All stable distributions are infinitely divisible."

17. "An infinitely divisible distribution must necessarily be n-divisible for every positive integer n."

18. Let $\{t_j\}$ (where j runs over the integers and $t_{j-1} < t_j$) denote a Poisson sequence of instants of time. That is, the random instants t_j are uncorrelated with each other, and the probability that r such instants occur in any time interval $(t_0, t_0 + t)$ is a Poisson distribution, given by $e^{-\lambda t}\,(\lambda t)^r / r!$. Let the random variable $n(0,t)$ denote the number of instants t_j that lie in the time interval $(0,t)$, and let $0 < t\,' < t$.

    "$n(0,t)$ and $n(0,t\,')$ are independent random variables."

19. *Continued*: "$\langle n^2(0,t)\rangle = \lambda t(\lambda t + 1)$."

20. *Continued*: "The autocorrelation $\langle n(0,t\,')\,n(0,t)\rangle$ is an exponentially decaying function of $(t - t\,')$."

21. Let $\xi(t)$ denote a stationary dichotomous Markov process that switches from the value c_1 to the value c_2 at a mean rate λ_1, and from c_2 to c_1 at a mean rate λ_2. Let $\delta\xi(t) = \xi(t) - \langle\xi\rangle$.

    "$\langle\delta\xi(t_1)\,\delta\xi(t_2)\rangle$ is an exponentially decaying function of $|t_1 - t_2|$ for all finite positive values of λ_1 and λ_2."

22. "If $X(t)$ is a continuous Markov process, all its joint probability densities can be expressed in terms of its probability density $p_1(x,t)$ and its two-time conditional probability density $p_2(x,t|x\,',t\,')$."

23. Consider a random walk in continuous time on an infinite linear lattice whose sites are labelled by the integer j. Let $P(j,t)$ be the probability that the walker is at the site j at time t, given that $P(j,0) = \delta_{j,0}$.

    "If the walk is unbiased, the probability $P(0,t)$ decays to zero like $t^{-1/2}$ as $t \to \infty$."

24. *Continued*: "If the walk is unbiased, the probability $P(j,t)$ for $j \neq 0$ decays to zero exponentially (apart from an inverse power of t) as $t \to \infty$."

25. *Continued*: "If the walk is biased, the probability $P(j,t)$ decays to zero exponentially (apart from an inverse power of t) as $t \to \infty$, for *all* j."

26. A particle undergoes unbiased diffusion on the infinite line $-\infty < x < \infty$, starting from $x = 0$ at $t = 0$. Its positional PDF $p(x,t)$ satisfies the diffusion equation $\partial p/\partial t = D\partial^2 p/\partial x^2$.

    "The mean time for the particle to reach the point $x = a$ (where $a > 0$) for the first time is proportional to a^2."

27. Let $\xi(t)$ satisfy the Langevin equation $\dot{\xi} = f(\xi) + g(\xi)\,\eta(t)$, where f and g are, in general, sufficiently smooth and differentiable functions of ξ. $\eta(t)$ is a Gaussian white noise, i.e., it is a Gaussian, stationary, Markov process with zero mean and autocorrelation $\langle \eta(t)\,\eta(t')\rangle = \delta(t - t')$.

    "Since f and g do not have any explicit dependence on t, it follows that $\xi(t)$ *must* be a stationary process."

28. *Continued*: "$\xi(t)$ is a Markov process if and only if f is at most a linear function of ξ, and g is a constant independent of ξ."

29. *Continued*: "For any f and g, $\xi(t)$ is a Markov process and its conditional PDF p satisfies the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial \xi}(fp) + \frac{1}{2}\frac{\partial^2}{\partial \xi^2}(g^2 p).''$$

30. *Continued*: "For any f and g, $\lim_{t\to\infty} p(\xi, t|\xi_0)$ always exists, and is given by

$$\lim_{t\to\infty} p(\xi, t|\xi_0) = p(\xi),$$

    where $p(\xi)$ is a normalizable stationary PDF."

31. *Continued*: "For any f and g, the autocorrelation function of $\xi(t)$ is of the form $\langle \xi(0)\,\xi(t)\rangle = \langle \xi^2\rangle\, e^{-\lambda t}$ where λ is a positive constant."

32. *Continued*: "For any f and g, the functional form of the conditional PDF of ξ is a Gaussian, since the noise η is a Gaussian random process."

33. "A one-dimensional, continuous, stationary Gaussian process whose autocorrelation function is an exponentially decaying function of time with a single relaxation time is *necessarily* the Ornstein-Uhlenbeck process."

34. "A one-dimensional, continuous, stationary Markov process whose autocorrelation function is an exponentially decaying function of time with a single relaxation time is *necessarily* a Gaussian process."

35. Consider a harmonically bound particle of mass m moving on the x-axis. The natural frequency is ω_0, and the damping constant is γ. In the high-friction approximation, the Langevin equation satisfied by the position $x(t)$ of the particle is given by

$$\dot{x} = -\frac{\omega_0^2}{\gamma}\, x + \Big(\frac{2k_B T}{m\gamma}\Big)^{1/2} \eta(t),$$

where $\eta(t)$ is a zero mean, δ-correlated Gaussian white noise.

"$x(t)$ is a stationary random variable, with a conditional PDF $p(x, t \,|\, x_0)$ that is given by an Ornstein-Uhlenbeck distribution."

36. *Continued*: The autocorrelation function $\langle x(0)\, x(t)\rangle$ decays exponentially to zero from an initial value $k_B\, T/(m\omega_0^2)$, with a relaxation time γ/ω_0^2."

37. A particle undergoing unbiased diffusion on the x-axis starts from the origin at $t = 0$. Its position $X(t)$ is given to be a Wiener process.

"The autocorrelation function $\langle X(t)\, X(t\,')\rangle$ is a function of the time difference $|t - t\,'|$."

38. *Continued*: "$X(t)$ is a continuous, Gaussian, Markov process whose sample paths are not differentiable almost everywhere."

39. *Continued*: "The mean number of times that $X(t)$ takes the value 0 in any finite interval of time $(0, T)$ is infinite."

40. Let $V(t)$ denote an Ornstein-Uhlenbeck process. Consider the process $X(t)$ whose increment is given by $dX(t) = V(t)\,dt$.

"$X(t)$ is a continuous, *non*-stationary, *non*-Markov process."

41. The power spectral density of stationary random process $X(t)$ is defined as

$$S_X(\omega) \stackrel{\text{def.}}{=} \lim_{T\to\infty} \left| \int_0^T dt\, e^{i\omega t}\, X(t) \right|^2.$$

"If the random process is ergodic, then $S_X(\omega)$ can be expressed in terms of an average over the realizations of $X(t)$ as

$$S_X(\omega) = (1/\pi) \int_0^\infty dt\, \langle X(0)\, X(t) \rangle \cos\, \omega t."$$

42. $f(t)$ is a pulse shape defined for $t \geq 0$, and $\{t_j\}$ is a Poisson sequence of instants of time with a mean separation $1/\nu$. Let $\{h_j\}$ be a set of *iidrv*. Consider the random process $\xi(t) = \sum_{j=-\infty}^{\infty} h_j\, f(t - t_j)$.

"According to the generalization of Campbell's Theorem, the r^{th} cumulant of $\xi(t)$ is given by

$$\kappa_r = \nu\, \langle h^r \rangle \int_{-\infty}^{\infty} dt\, |f(t)|^r$$

if and only if the pulse has a compact shape, i.e., if and only if $f(t)$ vanishes identically for t greater than some finite value τ."

43. *Continued*: "If $\langle \xi \rangle = 0$, the power spectral density $S_\xi(\omega)$ is proportional to $|\widetilde{f}(\omega)|^2$, where $\widetilde{f}(\omega)$ is the Fourier transform of $f(t)$."

44. Let $X(t)$ be a continuous, differentiable, stationary, Gaussian random process with zero mean and autocorrelation function $C(t) = \langle X(0)\, X(t) \rangle$. The variance of $X(t)$ is therefore $\sigma_X^2 = C(0)$.

"The random process $V(t) = \dot{X}(t)$ is also a stationary Gaussian process, with a variance given by $\sigma_V^2 = -\left[d^2C(t)/dt^2\right]_{t=0}$."

45. *Continued*: "The mean number of zero-crossings of $X(t)$ (i.e., the mean number of times that $X(t)$ takes the value 0) in a time interval $(0, T)$ is given by $\langle N(0; 0, T)\rangle = (\sigma_V T)/(\pi\sigma_X)$."

46. A particle undergoes dichotomous (or persistent) diffusion on the x-axis. Its velocity $V(t)$ is a stationary, symmetric dichotomous Markov process that switches randomly between the values c and $-c$ at a mean rate ν, where c is a positive constant. Let $p_{\rm R}(x,t)$ and $p_{\rm L}(x,t)$ denote, respectively, the joint position and velocity PDFs $p(x, V = c, t)$ and $p(x, V = -c, t)$. The initial conditions on these PDFs are given to be

$$p_{\rm R}(x,0) = p_{\rm L}(x,0) = \tfrac{1}{2}\,\delta(x).$$

"With the initial conditions as given above, $p_{\rm R}(-x,t) = p_{\rm L}(x,t)$."

47. *Continued*: The total positional PDF is $p(x,t) \stackrel{\rm def.}{=} p_{\rm R}(x,t) + p_{\rm L}(x,t)$.

"Although neither $p_{\rm R}(x,t)$ nor $p_{\rm L}(x,t)$ is a symmetric function of x, their sum $p(x,t)$ is a Gaussian in x, with a peak at $x = 0$ for all $t > 0$."

48. *Continued*: "For all $t > 0$, the mean squared displacement

$$\langle x^2(t)\rangle \stackrel{\rm def.}{=} \int_{-\infty}^{\infty} dx\, x^2\, p(x,t) = 2Dt,$$

where $D = c^2/(2\nu)$."

49. *Continued*: The speed of the particle undergoing dichotomous diffusion is always c, whether it is moving to the right or to the left on the x-axis.

"Hence the mean first-passage time to start at the origin and reach the point $a > 0$ for the first time is a *finite* quantity proportional to a/c."

50. Consider a simple random walk in discrete time on an infinite linear lattice whose sites are labelled by the integers. It is given that the probability of a step to the right is α, while that of a step to the left is $\beta = 1 - \alpha$, where $0 < \alpha < 1$. The random walk starts at some arbitrary site j at time $n = 0$.

"The total probability of the walker ever returning to the site j is equal to 1 only if $\alpha = \beta$."

51. *Continued*: "The total probability of the walker ever visiting the site $j-1$ is equal to 0 if $\alpha > \frac{1}{2}$."

52. *Continued*: "When $\alpha = \beta$, the probability that the walker returns to the site j for the first time at time $n = 4$ is given by $F(j, 4 \,|\, j) = \frac{1}{8}$."

53. *Continued*: "When $\alpha = \beta$, the probability $F(j, n \,|\, j)$ that the walker returns to the site j for the first time at time n decreases, for very large n, like $1/\sqrt{n}$."

54. Consider an unbiased continuous time random walk (CTRW) on a linear lattice, with a normalized waiting-time density given by $\psi(t) = 1/T$ for $0 \leq t \leq T$, and $\psi(t) = 0$ for $t > T$, where T is a positive constant.

    "The leading asymptotic behavior of the mean squared displacement in the long-time limit is $\sim t/T$."

55. *Continued*: If the waiting-time density $\psi(t)$ decays like $1/t^{1+\gamma}$ as $t \to \infty$, where $0 < \gamma < 1$, then the mean squared displacement $\sim t^{1+\gamma}$ in the long-time limit."

II. Fill in the blanks in the following.

1. A random variable X can take on the values 1 and -1, with respective probabilities α and $1-\alpha$, where $0<\alpha<1$. The characteristic function of X is $\widetilde{p}(k)=\cdots$

2. The moment generating function of the random variable X is given by $f(z)=\langle z^X\rangle$, where the average is over all realizations of the random variable. Let Y be the random variable $kX+l$, where k and l are given positive integers. In terms of f, the moment generating function of Y is $\phi(z)=\cdots$

3. Consider the set of n independent random variables $X_1,...,X_n$, where each X_j can take the value a (where $a>0$) with probability p_j, and the value b (where $|b|\leq a$) with probability $q_j=1-p_j$, with $0<p_j<1$.

   Let $Z_n=X_1+X_2+\cdots+X_n$. For a fixed value of a, the largest possible value of the variance of Z_n occurs when

   $$p_j=\cdots\,, \quad \text{and is given by} \quad \big[\mathrm{Var}\,(Z_n)\big]_{\max}=\cdots$$

4. Let m and n be two independent Poisson-distributed random variables with respective mean values μ and ν. Then, for any given positive integer l,

   $$\Pr\left(|m-n|=l\right)=\cdots$$

5. Consider a collection of *independent* particles in which each particle has a set of energy levels available to it. Let ε be one such energy level. The probability that a particle has energy ε is proportional to $e^{-\beta(\varepsilon-\mu)}$, where $\beta>0$ and $\mu<0$ are constants. There is no restriction on how many particles can occupy the energy level ε, i.e., the number of particles in the energy level ε can be $0,1,2,\ldots$ *ad infinitum*. The mean and variance of the number of particles in the level ε are then

   $$\langle n(\varepsilon)\rangle=\cdots \quad \text{and} \quad \mathrm{Var}\,[n(\varepsilon)]=\cdots$$

6. Let n be a random variable with a probability distribution given by the negative binomial distribution with mean μ and parameter N. This has the characteristic function

   $$\widetilde{p}(k)=\Big(\frac{N}{N+\mu-\mu\,e^{-ik}}\Big)^N.$$

   Therefore n is N-divisible into N *iidrv*, each of which has a geometric distribution. If m is one such component, the normalized probability distribution of m is given by $P_m=\cdots$

7. Let X_j $(j = 1, 2, \ldots, n)$ be *iidrv*, each with a uniform distribution in $[0, 1]$. Then the random variable

$$\xi = \lim_{n\to\infty} a_n \Big\{ \Big(\sum_{j=1}^{n} X_j \Big) - b_n \Big\}$$

has a normal distribution, where $a_n = \cdots$ and $b_n = \cdots$

8. Let X be a Gaussian random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$. Consider the random variable $Y = e^X$. The PDF of Y is $\rho(y) = \cdots$, and its mean and variance are $\langle Y \rangle = \cdots$ and $\mathrm{Var}\,(Y) = \cdots$

9. Let X and Y be independent Gaussian random variables, each with zero mean and unit variance. The PDF of the ratio $Z = X/Y$ is then given by $p(z) = \cdots$

10. Let X_j $(j = 1, 2, \ldots, n)$ be independent Gaussian random variables with $\langle X_j \rangle = \mu_j$ and $\mathrm{Var}\, X_j = \sigma_j^2$. The cumulant generating function of the random variable $Z_n = X_1 + X_2 + \cdots + X_n$ is then given by $K_{Z_n}(u) = \cdots$

11. Let X and Y be *iidrv*, each uniformly distributed in $[0, 1]$. The PDF of their product $Z = XY$ is given by

$$p(z) = \cdots, \quad \text{and} \quad \Pr\left(Z \le \tfrac{1}{2}\right) = \cdots$$

12. $X_1, X_2, \ldots, X_n$ are *iidrv* taking values in $(-\infty, \infty)$. The PDF of X_j is $p(x)$, and its cumulative distribution function (CDF) is

$$\Pr\,(X_j \le x) = \int_{-\infty}^{x} dx'\, p(x') \equiv F(x).$$

Let $M_n = \max\,(X_1, X_2, \ldots, X_n)$. In terms of $F(x)$ and $p(x)$, the CDF of M_n is $\cdots$ and its PDF is $\cdots$

13. *Continued*: Let $m_n = \min\,(X_1, X_2, \ldots, X_n)$. In terms of $F(x)$ and $p(x)$, the CDF of m_n is $\cdots$ and its PDF is $\cdots$

14. Let $\{t_j\}$ (where j runs over the integers and $t_{j-1} < t_j$) denote a Poisson sequence of instants of time: the random instants t_j are uncorrelated with each other, and the probability that r such instants occur in any time

interval $(t_0, t_0 + t)$ is a Poisson distribution, given by $e^{-\lambda t}\,(\lambda t)^r/r!$. Let the random variable $n(0,t)$ denote the number of instants t_j that lie in the time interval $(0,t)$, and let $\delta n(0,t) \equiv n(0,t) - \langle n(0,t)\rangle$. Then the first four moments of $\delta n(0,t)$ are

$$\langle \delta n(0,t)\rangle = \cdots\,,\ \langle [\delta n(0,t)]^2\rangle = \cdots\,,\ \langle [\delta n(0,t)]^3\rangle = \cdots\,,\ \langle [\delta n(0,t)]^4\rangle = \cdots$$

15. *Continued*: Let $0 < t\,' < t$. Then

$$\langle \delta n(0,t\,')\,\delta n(0,t)\rangle = \cdots\,,\ \text{ so that }\ \langle n(0,t\,')\,n(0,t)\rangle = \cdots$$

Further,

$$\langle [\delta n(0,t\,')]^2\,\delta n(0,t)\rangle = \cdots$$

16. Consider a symmetric dichotomous Markov process (DMP) that jumps between the values $+1$ and -1 at a mean rate ν. Let $n(t)$ be the number of jumps in the time interval $(0,t)$. The mean value of the random variable $X(t) = (-1)^{n(t)}$ is $\langle X(t)\rangle = \cdots$, and its variance is $\mathrm{Var}\,X(t) = \cdots$

17. *Continued*: The conditional probabilities corresponding to this symmetric DMP are given by

$$\begin{aligned} P(+,t_2\,|\,+,t_1) &= P(-,t_2\,|\,-,t_1) &&= e^{-\nu(t_2-t_1)}\,\cosh\,\nu(t_2-t_1) \\ P(+,t_2\,|\,-,t_1) &= P(-,t_2\,|\,+,t_1) &&= e^{-\nu(t_2-t_1)}\,\sinh\,\nu(t_2-t_1), \end{aligned}$$

where $+$ and $-$ denote $+1$ and -1, respectively, and $t_1 \le t_2$. The *joint* two-time probabilities are therefore given by

$$\begin{aligned} P(+,t_2\,;\,+,t_1) &= \cdots\,, \qquad P(-,t_2\,;\,-,t_1) &&= \cdots \\ P(+,t_2\,;\,-,t_1) &= \cdots\,, \qquad P(-,t_2\,;\,+,t_1) &&= \cdots \end{aligned}$$

18. *Continued*: Let $t_1 < t_2 < t_3$. The joint three-time probability

$$P(+,t_3\,;\,-,t_2\,;\,+,t_1) = \cdots$$

19. Let j be an integer-valued random variable corresponding to a birth-and-death process in discrete time n $(= 0, 1, 2, \ldots)$. At each time step, j either increases by 1 with a probability α, or decreases by 1 with a probability β, or remains at the same value with a probability γ, such that $\alpha + \beta + \gamma = 1$. Hence the probability $P(j,n)$ that the random variable has a value j at time n satisfies the recursion relation $P(j,n) = \cdots$

20. The population n of a certain animal species in a habitat is a random variable that can take on the values $0, 1, \ldots, N$. Let $P(n,t)$ denote the probability that there are n animals at time t. Taking into account the limited food supply and competition for food among the animals, it is found that the transition rate from n to $n+1$ is $\alpha_n = \lambda(N-n)$, and that from n to $n-1$ is $\beta_n = \mu n$, where λ and μ are positive constants with the dimensions of $[\text{time}]^{-1}$. (Thus there is no transition from 0 to -1, or from N to $N+1$, as required.) Assume that $n(t)$ is a Markov process. In terms of the explicit expressions given above for α_n and β_n, the master equation is

$$\frac{dP(n,t)}{dt} = \cdots \quad (1 \le n \le N-1).$$

21. *Continued*: The master equations satisfied by $P(0,t)$ and $P(N,t)$ are

$$\frac{dP(0,t)}{dt} = \cdots \quad \text{and} \quad \frac{dP(N,t)}{dt} = \cdots$$

22. *Continued*: As $t \to \infty$, the distribution $P(n,t)$ $(0 \le n \le N)$ tends to a stationary distribution that is a normalized binomial distribution, given by $P_{\text{st}}(n) = \cdots$

23. *Continued*: Hence the mean population in the stationary state is given by $\langle n \rangle_{\text{st}} = \cdots$, while its variance is given by $[\text{Var}\,(n)]_{\text{st}} = \cdots$

24. The master equation satisfied by the conditional PDF of a stationary, continuous Markov process is

$$\frac{\partial}{\partial t}\, p(x,t \,|\, x_0) = \int dx' \left\{ p(x',t \,|\, x_0)\, w(x \,|\, x') - p(x,t \,|\, x_0)\, w(x' \,|\, x) \right\}.$$

A simple model that is very useful in many physical situations is as follows: The transition rate $w(x \,|\, x')$ from x' to x is *independent* of the initial value x', and is a function of the *final* value x alone. Hence $w(x \,|\, x') = \lambda\, u(x)$, where λ is a constant with the dimensions of $1/[\text{time}]$. Using the fact that $\int dx'\, p(x',t \,|\, x_0) = 1$ (conservation of probability), the master equation then simplifies to $\partial p/\partial t = \cdots$

25. *Continued*: Taking the $t \to \infty$ limit of this equation, the left-hand side vanishes, while $p(x,t \,|\, x_0) \to p(x)$, the stationary PDF of X. It follows that $u(x)$ must be $p(x)$, apart from a normalization constant $\int dx\, u(x)$, which can be taken to be unity. Hence $u(x) = p(x)$. The problem reduces to solving the differential equation for $p(x,t \,|\, x_0)$ (e.g., by using

Laplace transforms), given the normalized PDF $p(x)$. The initial condition is $p(x, 0 \,|\, x_0) = \delta(x - x_0)$. The solution is $p(x, t \,|\, x_0) = \cdots$

The form of this solution is much simpler than that of the Ornstein-Uhlenbeck PDF. The Markov process $X(t)$ is called a **Kubo-Anderson process**.

26. *Continued*: To find the autocorrelation function of X, first find the conditional average for a given $X(0) = x_0$, according to

$$\overline{X(t)\, X(0)} \stackrel{\text{def.}}{=} \int dx\, x\, x_0\, p(x, t\,|x_0).$$

This gives $\overline{X(t)X(0)} = \cdots$.

Next, averaging over the initial value x_0 with the PDF $p(x_0)$, we have

$$\langle X(t)X(0)\rangle = \int dx_0\, \overline{X(t)\, X(0)}\, p(x_0).$$

Therefore $\langle X(t)X(0)\rangle = \cdots$

Subtracting out $\langle X\rangle^2 \stackrel{\text{def.}}{=} \int dx\, x^2\, p(x)$, we have

$$\big\langle \big(X(t) - \langle X\rangle\big)\big(X(0) - \langle X\rangle\big)\big\rangle = e^{-\lambda t}\, \mathrm{Var}\,(X),$$

showing that the Kubo-Anderson process is exponentially correlated.

27. *Continued*: It is clear that the Kubo-Anderson process is a generalisation of the dichotomous Markov process. A further generalization is the so-called **kangaroo process**. Here, instead of taking the transition probability per unit time to be $w(x \,|\, x') = \lambda\, u(x)$ where λ is a constant rate, one allows for an initial-state-dependent transition rate $\lambda(x')$. The transition probability per unit time then has the factorized form $w(x \,|\, x') = \lambda(x')\, u(x)$. Once again, this leads to a simplification of the master equation. We now have

$$\frac{\partial}{\partial t}\, p(x, t \,|\, x_0) = u(x) \int dx'\, \lambda(x')\, p(x', t \,|\, x_0) \; - \lambda(x)\, p(x, t \,|\, x_0) \int dx'\, u(x').$$

Passing to the limit $t \to \infty$, we find

$$u(x) = \frac{\lambda(x)}{\langle\lambda\rangle}\, p(x), \quad \text{where } \langle\lambda\rangle \stackrel{\text{def.}}{=} \int dx\, \lambda(x)\, p(x).$$

$\langle\lambda\rangle$ obviously represents the mean transition rate. Take the Laplace transform of the master equation (with respect to t). Define the function

$$\phi(s) = 1 - \frac{1}{\langle\lambda\rangle}\int dx\,\frac{\lambda^2(x)\,p(x)}{s+\lambda(x)} \equiv 1 - \frac{1}{\langle\lambda\rangle}\left\langle\frac{\lambda^2}{s+\lambda}\right\rangle.$$

The solution for the Laplace transform of the conditional PDF $p(x,t\,|\,x_0)$ is then found to be $\widetilde{p}(x,s\,|\,x_0) = \cdots$

28. *Continued*: The autocorrelation function corresponding to a kangaroo process is of interest in physical applications. Formally, we have

$$C(t) = \langle X(t)X(0)\rangle = \int dx_0 \int dx\, x_0\, x\, p(x,t\,|\,x_0)\, p(x_0).$$

The Laplace transform of the autocorrelation function is therefore

$$\widetilde{C}(s) = \int dx_0 \int dx\, x_0\, x\, \widetilde{p}(x,s\,|\,x_0)\, p(x_0).$$

Let the random variable X take values in $(-\infty,\infty)$, and assume that both $p(x)$ and $\lambda(x)$ are even functions of x. The autocorrelation function is then given by a superposition of decaying exponentials, given by $C(t) = \cdots$

29. A particle undergoes diffusion on the xy-plane, starting from the origin at $t = 0$. The normalized PDF of its position (X,Y) at any time $t > 0$ is given by

$$p(x,y,t) = (4\pi Dt)^{-1}\, e^{-(x^2+y^2)/(4Dt)},$$

where D is the diffusion constant. The normalized PDF of the random variable $U = X^2 + Y^2$ is $\rho(u,t) = \cdots$

30. Let v denote a Cartesian component of the velocity of a particle of mass m in a fluid in thermal equilibrium at temperature T. The autocorrelation of v is found to be $\langle v(0)v(t)\rangle = (k_BT/m)\,e^{-\gamma|t|}$ where γ is a positive dissipation constant. Hence the mean squared displacement of the particle at any time $t > 0$ is given by

$$\left\langle\big(x(t) - x(0)\big)^2\right\rangle = \cdots$$

31. The velocity autocorrelation function of a particle undergoing diffusion on the x-axis is found to be $\langle v(t_0)\,v(t_0+t)\rangle = a^2/(t^2+\tau^2)$, where a and τ are positive constants. The diffusion constant of the particle is therefore $D = \cdots$

32. The autocorrelation function of a continuous, stationary stochastic process $\xi(t)$ with zero mean is given to be $\langle \xi(0)\xi(t)\rangle = \langle \xi^2\rangle\, e^{-\alpha|t|}\cos\beta t$, where α and β are positive constants. The power spectral density of ξ,

$$S_\xi(\omega) \stackrel{\text{def.}}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, e^{i\omega t}\,\langle \xi(0)\xi(t)\rangle,$$

is therefore given by $S_\xi(\omega) = \cdots$

33. The velocity $\mathbf{v}$ of a particle of mass m in a fluid at temperature T satisfies the Langevin equation

$$m\dot{\mathbf{v}} = -m\gamma\mathbf{v} + (2m\gamma k_B T)^{1/2}\boldsymbol{\eta}(t),$$

where γ is the dissipation constant, and the noise $\boldsymbol{\eta}(t)$ is a vector-valued, stationary, Markov, Gaussian process. Its Cartesian components $\eta_j(t)$ satisfy

$$\langle \eta_j(t)\rangle = 0, \quad \langle \eta_j(t)\,\eta_k(t')\rangle = \delta_{kj}\,\delta(t-t') \quad (j,k=1,2,3).$$

Then the autocorrelation functions

$$\langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle = \cdots \quad \text{and} \quad \langle \mathbf{v}(0)\times\mathbf{v}(t)\rangle = \cdots$$

34. *Continued*: Suppose the particle has an electric charge q. In the presence of an applied magnetic field $\mathbf{B}$, the Langevin equation for its velocity becomes

$$m\dot{\mathbf{v}} = -m\gamma\mathbf{v} + q\,(\mathbf{v}\times\mathbf{B}) + (2m\gamma k_B T)^{1/2}\boldsymbol{\eta}(t).$$

The Fokker-Planck equation satisfied by the conditional probability density $p(\mathbf{v}, t|\mathbf{v}_0)$ is $\partial p/\partial t = \cdots$

In vector form, denoting by $\nabla_{\mathbf{v}}$ the gradient operator with respect to the components of $\mathbf{v}$, this equation is $\partial p/\partial t = \cdots$, which simplifies a bit to give $\partial p/\partial t = \cdots$

35. *Continued*: Let the magnetic field be directed along the unit vector $\mathbf{n}$ (so that $\mathbf{B} = B\mathbf{n}$), and let $\omega_{\text{c}} = qB/m$ denote the cyclotron frequency of the charged particle in the magnetic field. The autocorrelation of the velocity of the particle in the presence of the magnetic field is then given, for $t \geq 0$, by

$$C_{ij}(t) \stackrel{\text{def.}}{=} \langle v_i(0)\, v_j(t)\rangle = \frac{k_B T}{m}\, e^{-\gamma t}\left\{\cdots\right\}$$

Using the stationarity of the velocity process and the time-reversal properties of the reversible and irreversible parts of the drift terms in the Langevin

equation, it can be shown that the expression for $C_{ij}(t)$ for all t involves only the replacement of the factor $e^{-\gamma t}$ by $e^{-\gamma|t|}$ in the result found for $t > 0$. The terms in the curly brackets remain unchanged. Observe that the *symmetric* part of the tensor $C_{ij}(t)$ is an *even* function of t, while the *antisymmetric* part of the tensor is an *odd* function of t.

36. *Continued*: Hence, in the presence of a constant magnetic field, the autocorrelation functions

$$\langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle = \cdots \quad \text{and} \quad \langle \mathbf{v}(0) \times \mathbf{v}(t) \rangle = \cdots$$

37. *Continued*: The Kubo-Green formula for the diffusion tensor now becomes

$$D_{ij} = \tfrac{1}{2} \int_0^\infty dt \, [C_{ij}(t) + C_{ij}(-t)] = \tfrac{1}{2} \int_0^\infty dt \, [C_{ij}(t) + C_{ji}(t)].$$

Using the explicit result for $C_{ij}(t)$. the diffusion tensor is given by $D_{ij} = \cdots$

38. *Continued*: The expression for D_{ij} shows how the diffusion coefficient in the direction of the field (the longitudinal coefficient) differs from that in the plane normal to the field (the transverse coefficient). To see this explicitly, consider the case $\mathbf{n} = (0, 0, 1)$. The longitudinal and transverse and longitudinal diffusion coefficients are then given by

$$D_{\text{long}} = \cdots \quad \text{and} \quad D_{\text{trans}} = \cdots$$

The *effective* diffusion constant in the presence of a magnetic field can then be identified with the help of the asymptotic behaviour of the mean squared displacement, according to $\langle r^2(t) \rangle = \langle x^2(t) \rangle + \langle y^2(t) \rangle + \langle z^2(t) \rangle \sim 6 D_{\text{eft}} \, t$. This gives $D_{\text{eff}} = \cdots$

39. *Continued*: The dynamic mobility tensor of the particle is given by

$$\mu_{ij}(\omega) = \frac{1}{k_B T} \int_0^\infty dt \, e^{i\omega t} \, \langle v_i(0) \, v_j(t) \rangle.$$

Using the expression found above for $\langle v_i(0) \, v_j(t) \rangle$, the dynamic mobility in the presence of a magnetic field is given by $\mu_{ij}(\omega) = \cdots$

40. Let $X(t)$ denote the position of a particle undergoing free diffusion on the x-axis. Thus $X(t)$ is a Brownian motion, satisfying the stochastic differential equation (or Itô equation)

$$dX(t) = \sqrt{2D}\, dW(t).$$

Here $W(t)$ is a Wiener process, satisfying

$$\langle W(t) \rangle = 0, \quad \langle W(t)\, W(t') \rangle = \min(t, t').$$

The master equation for the PDF $p(x.t)$ of X is of course the diffusion equation,

$$\frac{\partial p}{\partial t} = D\, \frac{\partial^2 p}{\partial t^2}\,.$$

Let $\rho(\xi, t)$ denote the PDF of the random variable $\xi = X^n$, where n is a positive integer. It follows from the diffusion equation for $p(x, t)$ that $\rho(\xi, t)$ satisfies the master equation $\partial\rho/\partial t = \cdots$

41. *Continued*: Hence the stochastic differential equation (or Itô equation) satisfied by $\xi(t)$ is $d\xi(t) = \cdots$

42. A particle undergoes dichotomous (or persistent) diffusion on the x-axis. Its velocity $V(t)$ is a stationary, symmetric dichotomous Markov process that switches randomly between the values c and $-c$ at a mean rate ν, where c is a positive constant. It is convenient to write the joint PDF of the position X and velocity V as

$$p(x, V = c, t) \equiv p_{\mathrm{R}}(x, t) \quad \text{and} \quad p(x, V = -c, t) \equiv p_{\mathrm{L}}(x, t),$$

the subscripts R and L standing for 'right-moving' and 'left-moving', respectively. These PDFs satisfy the coupled partial differential equations

$$\Big(\frac{\partial}{\partial t} + c\, \frac{\partial}{\partial x}\Big)\, p_{\mathrm{R}} = \nu\, (p_{\mathrm{L}} - p_{\mathrm{R}}) \ \text{ and } \ \Big(\frac{\partial}{\partial t} - c\, \frac{\partial}{\partial x}\Big)\, p_{\mathrm{L}} = \nu\, (p_{\mathrm{R}} - p_{\mathrm{L}}).$$

The positional PDF of the particle is

$$p(x, t) \stackrel{\text{def.}}{=} p_{\mathrm{R}}(x, t) + p_{\mathrm{L}}(x, t).$$

Now consider the case in which the particle starts at $t = 0$ with position $X = 0$ and velocity $V = c$, i.e., it *starts in the right-moving state*. The initial conditions on the PDFs $p_{\mathrm{R}}(x, t)$ and $p_{\mathrm{L}}(x, t)$ are therefore

$$p_{\mathrm{R}}(x, 0) = \delta(x) \quad \text{and} \quad p_{\mathrm{L}}(x, 0) = 0.$$

It follows that the initial conditions on their time derivatives are given by

$$\Big[\frac{\partial p_{\mathrm{R}}(x,t)}{\partial t}\Big]_{t=0} = \cdots \quad \text{and} \quad \Big[\frac{\partial p_{\mathrm{L}}(x,t)}{\partial t}\Big]_{t=0} = \cdots$$

Hence the initial conditions on the positional PDF $p(x,t)$ and its time derivative are

$$p(x,0) = \cdots \quad \text{and} \quad \Big[\frac{\partial p(x,t)}{\partial t}\Big]_{t=0} = \cdots$$

43. *Continued*: From the coupled differential equations for p_{R} and p_{L}, it is easily shown that each of the PDFs $p_{\mathrm{R}}(x,t)$, $p_{\mathrm{L}}(x,t)$ and $p(x,t)$ satisfies the second-order partial differential equation

$$\Big(\frac{\partial^2}{\partial x^2} + 2\nu\,\frac{\partial}{\partial t} - c^2\,\frac{\partial^2}{\partial x^2}\Big)\,f(x,t) = 0,$$

where f stands for p_{R}, or p_{L}, or p. The mean displacement and mean squared displacement at any time $t \geq 0$ are, by definition, the first and second moments

$$\langle X(t)\rangle \stackrel{\text{def.}}{=} \int_{-\infty}^{\infty} dx\, x\, p(x,t) \quad \text{and} \quad \langle X^2(t)\rangle \stackrel{\text{def.}}{=} \int_{-\infty}^{\infty} dx\, x^2\, p(x,t).$$

It follows from the differential equation for $p(x,t)$ that these moments satisfy the second-order ordinary differential equations

$$\frac{d^2}{dt^2}\langle X(t)\rangle = \cdots \quad \text{and} \quad \frac{d^2}{dt^2}\langle X^2(t)\rangle = \cdots$$

44. From the initial conditions on $p(x,t)$ and $\partial p(x,t)/\partial t$, it follows that the initial conditions satisfied by $\langle X(t)\rangle$ and $\langle X^2(t)\rangle$ are

$$\langle X(0)\rangle = \cdots \quad \text{and} \quad \Big[\frac{d\langle X(t)\rangle}{dt}\Big]_{t=0} = \cdots$$

and

$$\langle X^2(0)\rangle = \cdots \quad \text{and} \quad \Big[\frac{d\langle X^2(t)\rangle}{dt}\Big]_{t=0} = \cdots$$

Hence the solutions for the first and second moments of $X(t)$ are

$$\langle X(t)\rangle = \cdots \quad \text{and} \quad \langle X^2(t)\rangle = \cdots$$

The variance of $X(t)$ is therefore $\operatorname{Var} X(t) = \cdots$

45. *Continued*: Let the random variable $N(0;0,T)$ denote the number of crossings of the origin (the point $x = 0$) by the particle in the time interval $(0,T)$ in the case of dichotomous diffusion, with the initial conditions already specified. The mean value of $N(0;0,T)$ can then be shown to reduce to the expression

$$\langle N(0;0,T)\rangle = \tfrac{1}{2}\nu \int_0^T dt\, e^{-\nu t}\,[I_0(\nu t) + I_1(\nu t)] = \tfrac{1}{2}\int_0^{\nu T} du\, e^{-u}\,[I_0(u) + I_1(u)],$$

where I_j is the modified Bessel function of the first kind and of order j. Using the asymptotic form of $I_j(u)$ as $u \to \infty$, it follows that the leading large-T behaviour of the mean number of zero-crossings in $(0,T)$ is given by $\langle N(0;0,T)\rangle \sim \cdots$

46. *Continued*: Further, the *variance* of $N(0;0,T)$ can be shown to reduce to the expression

$$\mathrm{Var}\,[N(0;0,T)] = \nu T - 2\langle N(0;0,T)\rangle.$$

The standard deviation $\Delta N(0;0,T)$ is of course the square root of the variance of $N(0;0,T)$. It follows that the relative fluctuation in $N(0;0,T)$ tends to the limiting value

$$\lim_{T\to\infty} \frac{\Delta N(0;0,T)}{\langle N(0;0,T)\rangle} = \cdots$$

47. Consider a finite linear lattice comprising the sites $0, 1, 2, \ldots, j$. A random walker executes a simple unbiased random walk in discrete time on this lattice, by making nearest-neighbor jumps from any site to its neighbouring sites at the end of every time step. The random walk is a Markov chain. Let T_k denote the mean time for the walker to reach the end-point j for the first time, starting from any site k. It is obvious that $T_j = 0$ by definition.

The recursion relation satisfied by T_k for $1 \le k \le j-2$ is $T_k = \cdots$, while those satisfied by T_0 and T_{j-1} are $T_0 = \cdots$ and $T_{j-1} = \cdots$

The solution for T_k (where $0 \le k \le j$) is $T_k = \cdots$

48. Consider an unbiased simple random walk in discrete time on an infinite linear lattice whose sites are labelled by the integers. The random walk

starts at time $n = 0$ from the site 0. The probability that the walker is at any site j at time n is given by the binomial distribution

$$P(j, n \,|\, 0) = \begin{cases} {}^{n}C_{(n-j)/2}\; 2^{-n}, & \text{if } |j| \leq n \text{ and } n - j \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

$P(0, n \,|\, 0)$ is the probability of a return (not necessarily the first return!) to the origin at time n. The generating function of this distribution is

$$\pi_{00}(z) \stackrel{\text{def.}}{=} \sum_{n=1}^{\infty} P(0, n \,|\, 0)\, z^n = \cdots$$

$\Big[$You will need the formula $\sum\limits_{k=1}^{\infty} {}^{2k}C_k \left(\frac{1}{2}z\right)^{2k} = \dfrac{1 - \sqrt{1 - z^2}}{\sqrt{1 - z^2}}$.$\Big]$

Now let $F(0, n \,|\, 0)$ be the probability that the *first* return to the origin occurs at time n. The corresponding generating function is

$$\phi_{00}(z) \stackrel{\text{def.}}{=} \sum_{n=1}^{\infty} F(0, n \,|\, 0)\, z^n.$$

From the result found above for $\pi_{00}(z)$, it follows that $\phi_{00}(z) = \cdots$

Hence the mean first-return time or *mean recurrence time* to the origin is $\langle t(0 \,|\, 0) \rangle = \cdots$

On an infinite lattice, owing to translational invariance, $\pi_{00}(z) = \pi_{jj}(z)$, and hence $\phi_{00}(z) = \phi_{jj}(z)$, for any site j.

49. *Continued*: $P(1, n \,|\, 0)$ is the probability that the walker, having started at site 0, is at the neighbouring site 1 at time n. The generating function of $P(1, n \,|\, 0)$ is

$$\pi_{10}(z) \stackrel{\text{def.}}{=} \sum_{n=1}^{\infty} P(1, n \,|\, 0)\, z^n = \cdots$$

$\Big[$You will need the formula $\sum\limits_{k=1}^{\infty} {}^{2k-1}C_{k-1} \left(\frac{1}{2}z\right)^{2k-1} = \dfrac{1 - \sqrt{1 - z^2}}{z\sqrt{1 - z^2}}$.$\Big]$

Now consider the probability $F(1, n \,|\, 0)$ that the walker is at the site 1 for the *first time* at time n. The corresponding generating function of this distribution is given by

$$\phi_{10}(z) \stackrel{\text{def.}}{=} \sum_{n=1}^{\infty} F(1, n \,|\, 0)\, z^n.$$

It follows from the results for $\pi_{10}(z)$ and $\pi_{11}(z)$ that $\phi_{10}(z)$ is given by $\phi_{10}(z) = \cdots$

Hence the *mean first-passage time* to go from the site 0 to the site 1 is $\langle t(1|0)\rangle = \cdots$

50. *Continued*: The recurrence-time distribution $F(0, n \,|\, 0) = \cdots$ and the first-passage-time distribution $F(1, n \,|\, 0) = \cdots$

51. The **backward Kolmogorov equation** is a very useful tool in the determination of first-passage times in the case of Markov chains and Markov processes. Here is a general formulation applicable to Markov chains, and hence to discrete-time Markovian random walks on a lattice (or a graph, more generally). Let $j, k, l, \ldots$ label the states of a Markov chain (or vertices of a graph, or 'sites'). For simplicity let's consider a finite chain and a transition matrix that connects each site k to a set of specified sites that we shall call the 'nearest neighbors' of the site k. Let ν_k be the number of such sites, i.e., the coordination number of the site k. A jump (or a one-time-step transition) can from each site k to any one of its nearest-neighbor (nn) sites with probability $1/\nu_k$. There is also some specified set of sites $\{\alpha\}$ that are traps: the Markov chain (or random walk) ends when any one of the trap sites is reached. The task is to determine the probability distribution $F(n, k)$ of the time-to-trapping t_k from any starting site k, and hence the mean time-to-trapping. $F(n, k)$ is the probability that, starting from k, the random walk hits a trap site at time n. It is the first-passage time (or hitting time, or trapping time) distribution. By definition, $F(n, \alpha) = \delta_{n,0}$ for each of the trap sites.

We can show that $F(n, k)$ satisfies the discrete version of the backward Kolmogorov equation, as follows. The first jump from k can only occur to one of its nearest-neighbor sites (with probability $1/\nu_k$). It is therefore clear that, for $n \geq 0$,

$$F(n+1, k) = \frac{1}{\nu_k} \sum_l F(n, l) \;\text{ where } l \text{ is a nearest-neighbour of } k,$$

because after the first time step, n time steps are left. Let us introduce the notation $\langle kl\rangle$ to indicate that l is a nearest neighbor of k, and $\delta_{\langle kl\rangle}$ to denote the corresponding Kronecker delta (= 1 if l is a nearest neighbor of k, and = 0 if it is not). Using the summation convention for repeated indices,

$$F(n+1, k) = \frac{1}{\nu_k} \delta_{\langle kl\rangle}\, F(n, l), \quad \text{for} \quad n \geq 0.$$

Therefore

$$F(n+1,k) - F(n,k) = \left\{ \frac{1}{\nu_k} \delta_{\langle kl \rangle} - \delta_{kl} \right\} F(n,l) \equiv \Delta_{kl}\, F(n,l).$$

The quantity in curly brackets, Δ_{kl}, is the discrete analog of the Laplacian operator at the site k. The equation above is the discrete analog of the backward Kolmogorov equation for the first-passage-time distribution $F(n,k)$ on the Markov chain (or graph).

The general q^{th} moment of the first-passage time (or time-to-trapping) is of course defined as

$$T_k^{(q)} = \sum_{n=0}^{\infty} n^q\, F(n,k) = \sum_{n=1}^{\infty} n^q\, F(n,k),$$

because $F(0,k) \equiv 0$ as long as k is not a trap site α. For a trap site α, of course, we have $T_\alpha^{(q)} \equiv 0$ for every $q \geq 1$. Since first passage from any site k to a trap site α is a sure event, we have

$$T_k^{(0)} = \sum_{n=0}^{\infty} F(n,k) = \sum_{n=1}^{\infty} F(n,k) = 1.$$

The quantity of primary interest is the **mean first-passage time**, or mean time-to-trapping starting from any site k, given by the first moment of $F(n,k)$, i.e.,

$$T_k^{(1)} \equiv \langle t_k \rangle = \sum_{n=0}^{\infty} n\, F(n,k).$$

This quantity satisfies the equation $\Delta_{kl}\, T_l^{(1)} = \cdots$

52. *Continued*: Multiplying the equation $\Delta_{kl}\, F(n,l) = F(n+1,k) - F(n,k)$ by n^q (where q is any positive integer) and summing over n, the q^{th} moment of the time to trapping satisfies the equation $\Delta_{kl}\, T_l^{(q)} = \cdots$. In particular, the second moment of the time-to-trapping satisfies the equation $\Delta_{kl}\, T_l^{(2)} = \cdots$.

53. *Continued*: Let's apply the foregoing to a simple unbiased random walk on a finite linear lattice whose sites are labelled by the integers from $-j$ to j, these two end points being traps. Let $T_k^{(1)}$ be the mean time for a walker starting from any site k (where $-(j-1) \leq k \leq (j-1)$) to reach a trap site for the first time. The backward Kolmogorov equation for the mean first-passage time becomes $\cdots = \cdots$, showing that the second difference

of $T_k^{(1)}$ (with respect to k) is a constant. Hence $T_k^{(1)}$ must be a quadratic function of k. Further, it has the obvious symmetry property $T_k^{(1)} = T_{-k}^{(1)}$, because the traps are situated symmetrically on either side of the site 0, at $\pm j$. Finally, $T_k^{(1)}$ satisfies the boundary conditions $T_{\pm j}^{(1)} = 0$. Using these facts, the explicit solution for the mean time-to-trapping is $T_k^{(1)} = \cdots$

54. *Continued*: Continuing with the random walk on a linear lattice considered in the preceding question, the difference equation satisfied by the second moment of the time-to-trapping is

$$\tfrac{1}{2}\big(T_{k-1}^{(2)} + T_{k+1}^{(2)}\big) - T_k^{(2)} = \cdots .$$

Thus, the second difference of the quantity $T_k^{(2)}$ is a quadratic function of k. Hence the general solution for $T_k^{(2)}$ must be a quartic function of k. Using the symmetry property $T_k^{(2)} = T_{-k}^{(2)}$ and the boundary condition $T_{\pm j}^{(2)} = 0$, the explicit solution for the mean squared time-to-trapping is $T_k^{(2)} = \cdots$

55. *Continued*: The variance of the time-to-trapping t_k from the site k is therefore $\mathrm{Var}\,(t_k) = \cdots$. Hence the relative fluctuation, defined as the ratio of the standard deviation of t_k to its mean value, is given by $\Delta t_k / \langle t_k \rangle = \cdots$

56. Consider an unbiased continuous time random walk (CTRW) on an infinite linear lattice whose sites are labelled by the integer $j \in \mathbb{Z}$. The walk starts from the site 0 at $t = 0$. The normalized waiting time density of the time between successive jumps is $\psi(t)$, and the mean time between jumps is

$$\tau \stackrel{\text{def.}}{=} \int_0^\infty dt\, t\, \psi(t) = -\big[d\widetilde{\psi}(s)/ds\big]_{s=0} ,$$

in terms of the Laplace transform $\widetilde{\psi}(s)$ of $\psi(t)$. When $\psi(t)$ has any functional form other than an exponentially decaying function of t, the random walk becomes a non-Markovian process. That is, a CTRW is Markovian if and only if $\psi(t)$ is of the form $\lambda\, e^{-\lambda t}$.

Let $W(n, t)$ be the probability that exactly n jumps occur in the time interval t. Its Laplace transform is given by $\widetilde{W}(n, s) = \cdots$

57. *Continued*: The random walk is unbiased, so that the probability of a jump from site j to $j + 1$ or to $j - 1$ is equal to $\frac{1}{2}$. The generating function for

a single step is therefore $g(z) = \frac{1}{2}\,(z + z^{-1})$, and that for n steps is $[g(z)]^n$. That is, the probability $P_n(j)$ that the walker is at j after n steps is the coefficient of z^j in the expansion of $[g(z)]^n$ in powers of z. Equivalently,

$$\sum_{j\in\mathbb{Z}} P_n(j)\, z^j = [g(z)]^n.$$

Turning to continuous time, the probability $P(j,t)$ that the walker is at the site j at time t is therefore

$$P(j,t) = \sum_{n=0}^{\infty} P_n(j)\, W(n,t), \quad \text{so that} \quad \widetilde{P}(j,s) = \sum_{n=0}^{\infty} P_n(j)\, \widetilde{W}(n,s).$$

Let

$$L(z,t) = \sum_{j\in\mathbb{Z}} P(j,t)\, z^j$$

be the generating function of the probability distribution $P(j,t)$. Combining the results in the foregoing, its Laplace transform $\widetilde{L}(z,s)$ is given by the explicit expression $\widetilde{L}(z,s) = \cdots$

58. *Continued*: The Laplace transform of the mean displacement in time t is given by

$$\mathcal{L}[\langle j(t)\rangle] = \left[\partial\widetilde{L}(z,s)/\partial z\right]_{z=1} = \cdots .$$

The Laplace transform of the mean squared displacement in time t is therefore given by $\mathcal{L}[\langle j^2(t)\rangle] = \cdots$

59. The leading asymptotic ($t \to \infty$) behaviour of the mean squared displacement is therefore given by $\langle j^2(t)\rangle \sim \cdots$

60. *Continued*: Suppose the Laplace transform of the normalized waiting-time density has a small-s behaviour given by $\widetilde{\psi}(s) = 1 - cs^\gamma +$ higher powers of s. The leading long-time behavior of the mean squared displacement is then given by $\langle j^2(t)\rangle \sim \cdots$

61. Consider the Bernoulli shift map of the unit interval, given by the recursion relation $x_{n+1} = f(x_n) = 2x_n \mod 1$, where $n = 0, 1, 2, \ldots$ and $x_0 \in [0,1]$. Let the unit interval be partitioned into two cells $L = [0, \frac{1}{2}]$ and $R = (\frac{1}{2}, 1]$. The normalized invariant density of the map is given to be $\rho(x) = 1$. The

invariant measures of the two cells are therefore given by $\mu_L = \int_L dx = \frac{1}{2}$ and $\mu_R = \int_R dx = \frac{1}{2}$. We are interested in the statistics of recurrences to the cell R, say. Let

$$\widetilde{w}_n \stackrel{\text{def.}}{=} P(L, n-1; L, n-2; \cdots; L, 1; L, 0)$$

be the joint probability that the system starts in L (the complement of R) at time 0 and remains in L at times $1, 2, \ldots, n-1$. Then $\widetilde{w}_n$ is given by the multiple integral $\widetilde{w}_n = \cdots$.

62. *Continued*: Evaluating the integral, the explicit expression for this joint probability is $\widetilde{w}_n = \cdots$. Hence the normalized sojourn probability in L is $\widetilde{H}_n = \cdots$

63. *Continued*: The escape time probability distribution of the time of escape out of L is defined as the conditional probability

$$\widetilde{E}_n \stackrel{\text{def.}}{=} P(R, n; L, n-1; \cdots; L, 1 | L, 0).$$

In terms of the quantities $\{\widetilde{w}_n\}$, this is given by $\widetilde{E}_n = \cdots$. The final result for this quantity is $\widetilde{E}_n = \cdots$.

64. *Continued*: The distribution of the time of recurrence to R is defined as the conditional probability

$$R_n \stackrel{\text{def.}}{=} P(R, n; L, n-1; \cdots; L, 1 | R, 0).$$

In terms of the quantities $\{\widetilde{w}_n\}$, this is given by $R_n = \cdots$. The final result for the recurrence time distribution is $R_n = \cdots$. Hence the mean time of recurrence to R is $\cdots$

65. *Continued*: The partitioning of the unit interval into the cells L and R in the Bernoulli shift map is a Markov partition, in the following sense: The dynamics of transitions between the cells L and R turns out to be just that of a two-state Markov chain with the one-step conditional probabilities $P(L, 1|L, 0) = \ldots$, $P(L, 1|R, 0) = \cdots$, $P(R, 1|L, 0) = \cdots$ and $P(L, 1|L, 0) = \cdots$ playing the role of the transition probabilities.

66. *Continued*: As already stated, the recurrence time distribution for the cell R is defined as the conditional probability $P(R, n; L, n-1; \cdots; L, 1|R, 0)$. Evaluating this quantity under the assumption that the process is a Markov chain with transition probabilities as found above, the expression obtained for R_n is $\cdots$

Quiz: Solutions

I. True or false:

1. True
2. False
3. False
4. False
5. False
6. True
7. True
8. False
9. True
10. False
11. True
12. True
13. True
14. True
15. False
16. True
17. True
18. False
19. True
20. False
21. True
22. True
23. True
24. False

25. True
26. False
27. False
28. False
29. True
30. False
31. False
32. False
33. True
34. False
35. True
36. True
37. False
38. True
39. True
40. True
41. True
42. False
43. True
44. True
45. True
46. True
47. False
48. False
49. False
50. True

51. False

52. True

53. False

54. True

55. False

II. Fill in the blanks:

1. $\widetilde{p}(k) = \alpha\, e^{-ik} + (1-\alpha)\, e^{ik}$.

2. $\phi(z) = \langle z^Y \rangle = z^l\, f(z^k)$.

3. $p_j = \frac{1}{2}\,, \quad \big[\mathrm{Var}\,(Z_n)\big]_{\max} = na^2$.
.

4. $\Pr\,(|m-n| = l) = e^{-(\mu+\nu)}[(\mu/\nu)^{l/2} + (\nu/\mu)^{l/2}]\, I_l(2\sqrt{\mu\nu})$.

5. $\langle n(\varepsilon)\rangle = \dfrac{1}{e^{\beta(\varepsilon-\mu)}-1}, \quad \mathrm{Var}\,[n(\varepsilon)] = \dfrac{e^{\beta(\varepsilon-\mu)}}{[e^{\beta(\varepsilon-\mu)}-1]^2}$.

6. $P_m = \dfrac{N}{(N+\mu)}\Big(\dfrac{\mu}{N+\mu}\Big)^m$.

7. $a_n = \sqrt{12/n} \quad \text{and} \quad b_n = n/2$.

8. $\rho(y) = \dfrac{1}{y\sqrt{2\pi}}\, e^{-\frac{1}{2}(\ln\, y)^2}, \;\; \langle Y\rangle = \sqrt{e}\,, \;\; \mathrm{Var}\,(Y) = e(e-1)$.

9. $p(z) = \dfrac{1}{\pi}\, \dfrac{1}{(x^2+1)}$.

10. $K_{Z_n}(u) = \sum\limits_{j=1}^{n} \big[u\mu_j + \frac{1}{2}u^2\sigma_j^2\big]$.

11. $p(z) = -\ln\, z, \quad \Pr\,\big(Z \le \frac{1}{2}\big) = \frac{1}{2}(1 + \ln\, 2)$.

12. The CDF of M_n is $[F(x)]^n$. Its PDF is $n[F(x)]^{n-1}\,F'(x) = n[F(x)]^{n-1}\,p(x)$.

13. The CDF of m_n is $1 - [1 - F(x)]^n$. Its PDF is $n[1 - F(x)]^{n-1}\,p(x)$.

14. Since the instants $\{t_j\}$ are Poisson-distributed, it follows that the mean value of the random variable $n(0,t)$ is just $\langle n(0,t)\rangle = \lambda t$. Moreover, all its cumulants are also equal to λt. Hence

$$\langle \delta n(0,t)\rangle \equiv 0, \quad \langle [\delta n(0,t)]^2\rangle = \langle [\delta n(0,t)]^3\rangle = \lambda t.$$

Further,

$$\langle [\delta n(0,t)]^4\rangle - 3\langle [\delta n(0,t)]^2\rangle^2 = \lambda t,$$

so that

$$\langle [\delta n(0,t)]^4\rangle = \lambda t + 3\lambda^2 t^2.$$

15. Since $(0,t')$ and (t',t) are non-overlapping intervals, $n(0,t')$ and $n(t',t)$ are independent random variables, and hence so are $\delta n(0,t')$ and $\delta n(t',t)$. Therefore

$$\langle \delta n(0,t')\,\delta n(t',t)\rangle = \langle \delta n(0,t')\rangle\,\langle \delta n(t',t)\rangle \equiv 0.$$

But note that, when $t' < t$, we have

$$n(t',t) = n(0,t) - n(0,t'), \text{ so that } \delta n(t',t) = \delta n(0,t) - \delta n(0,t').$$

It follows that $\langle \delta n(0,t')\,\delta n(0,t)\rangle - \langle [\delta n(0,t')]^2\rangle = 0$, so that

$$\langle \delta n(0,t')\,\delta n(0,t)\rangle = \lambda t'.$$

Therefore

$$\langle n(0,t')\,n(0,t)\rangle = \lambda t' + \lambda^2\,t'\,t.$$

(More generally, $\langle \delta n(0,t')\,\delta n(0,t)\rangle = \lambda\,\min\,(t',t)$.)

Similarly, using the fact that

$$\langle [\delta n(0,t')]^2\,\delta n(t',t)\rangle = \langle [\delta n(0,t')]^2\rangle\,\langle \delta n(t',t)\rangle \equiv 0,$$

it follows that

$$\langle [\delta n(0,t')]^2\,\delta n(0,t)\rangle = \langle [\delta n(0,t')]^3\rangle = \lambda t'.$$

16. Since $(-1)^n = \pm 1$ according as n is even or odd, we have

$$\begin{aligned}\langle X(t)\rangle &= (+1)\times \Pr\{n(t)\,\text{is even}\} + (-1)\times \Pr\{n(t)\,\text{is odd}\}\\ &= e^{-\nu t}\,(\cosh\,\nu t - \sinh\,\nu t) = e^{-2\nu t}.\end{aligned}$$

Since $X^2(t)\equiv 1$, we have $\operatorname{Var} X(t) = 1 - e^{-4\nu t}$.

17. The DMP is stationary, so that $P(\pm, t_1) = P(\pm) = \frac{1}{2}$. Hence the joint two-time probabilities are

$$\begin{aligned}P(+,t_2\,;+,t_1) &= P(-,t_2\,;-,t_1) &&= \tfrac{1}{2}\,e^{-\nu(t_2-t_1)}\cosh\,\nu(t_2-t_1)\\ P(+,t_2\,;-,t_1) &= P(-,t_2\,;+,t_1) &&= \tfrac{1}{2}\,e^{-\nu(t_2-t_1)}\sinh\,\nu(t_2-t_1).\end{aligned}$$

18. Using the Markov property of the process, the joint three-time probability

$$P(+,t_3\,;-,t_2\,;+,t_1) = \tfrac{1}{2}\,e^{-\nu(t_3-t_1)}\sinh\,\nu(t_3-t_2)\sinh\,\nu(t_2-t_1).$$

19. $P(j,n) = \alpha\,P(j-1,n-1) + \beta\,P(j+1,n-1) + \gamma\,P(j,n-1)$.

20. For $1 \le n \le N-1$,

$$\frac{dP(n,t)}{dt} = \lambda(N{-}n{+}1)P(n{-}1,t){+}\mu(n{+}1)P(n{+}1,t){-}[\lambda(N{-}n){+}\mu n]P(n,t).$$

21. $\dfrac{dP(0,t)}{dt} = -\lambda N P(0,t) + \mu P(1,t)$, $\dfrac{dP(N,t)}{dt} = \lambda P(N-1,t) - \mu N P(N,t)$.

22. $P_{\rm st}(n) = \left(\dfrac{\mu}{\lambda+\mu}\right)^N \binom{N}{n}\left(\dfrac{\lambda}{\mu}\right)^n$.

23. $\langle n\rangle_{\rm st} = \dfrac{N\lambda}{\lambda+\mu}$, $[\operatorname{Var}(n)]_{\rm st} = \dfrac{N\lambda\mu}{(\lambda+\mu)^2}$.

24. $\dfrac{\partial}{\partial t}\,p(x,t|x_0) = \lambda\,u(x) - \lambda\,p(x,t|x_0)\displaystyle\int dx'\,u(x')$.

25. $p(x,t\,|\,x_0) = \delta(x-x_0)\,e^{-\lambda t} + p(x)\,(1-e^{-\lambda t})$.

26. $\overline{X(t)X(0)} = x_0^2\,e^{-\lambda t} + x_0\langle X\rangle\,(1-e^{-\lambda t})$, where $\langle X\rangle = \int dx\,x\,p(x)$.

$\langle X(t)X(0)\rangle = \langle X^2\rangle\,e^{-\lambda t} + \langle X\rangle^2\,(1-e^{-\lambda t})$, where $\langle X^2\rangle = \int dx\,x^2\,p(x)$.

27. $\widetilde{p}(x,s\,|\,x_0) = \dfrac{\delta(x-x_0)}{s+\lambda(x)} + \dfrac{p(x)\,\lambda(x)\,\lambda(x_0)}{\langle\lambda\rangle\big(s+\lambda(x)\big)\big(s+\lambda(x_0)\big)\,\phi(s)}$.

When $\lambda(x) = \lambda$, a constant, the inverse Laplace transform is easily obtained. It is just the conditional PDF of the Kubo-Anderson process found earlier.

28. $C(t) = \displaystyle\int_{-\infty}^{\infty} dx\,e^{-\lambda(x)\,t}\,x^2\,p(x)$.

$C(t)$ is therefore given by a continuous superposition of decaying exponentials in time. As a consequence, $C(t)$ can exhibit diverse kinds of asymptotic behaviour as $t\to\infty$. These include power-law decay, stretched-exponential decay and exponential decay with a spectrum of relaxation times.

29. $\rho(u,t) = (4Dt)^{-1}e^{-u/(4Dt)}$.

30. $\langle\big(x(t)-x(0)\big)^2\rangle = \dfrac{k_BT}{m\gamma^2}\,\big(\gamma t - 1 + e^{-\gamma t}\big)$.

31. $D = \dfrac{\pi a^2}{2\tau}$.

32. $S_\xi(\omega) = \dfrac{\alpha\langle\xi^2\rangle}{2\pi}\left\{\dfrac{1}{\alpha^2+(\omega+\beta)^2} + \dfrac{1}{\alpha^2+(\omega-\beta)^2}\right\}$.

33. $\langle\mathbf{v}(0)\cdot\mathbf{v}(t)\rangle = \dfrac{3k_BT}{m}\,e^{-\gamma|t|}$, $\quad\langle\mathbf{v}(0)\times\mathbf{v}(t)\rangle = 0$.

34. $\dfrac{\partial p}{\partial t} = \dfrac{\partial}{\partial v_j}\Big[\Big(\gamma v_j - \dfrac{q}{m}\epsilon_{jkl}v_kB_l\Big)p\Big] + \dfrac{\gamma k_BT}{m}\delta_{jk}\dfrac{\partial^2 p}{\partial v_j\partial v_k}$.

In vector form, with $\nabla_{\mathbf{v}}$ standing for the gradient operator with respect to the components of $\mathbf{v}$,

$$\partial p/\partial t = \gamma\nabla_{\mathbf{v}}\cdot(\mathbf{v}p) - (q/m)\nabla_{\mathbf{v}}\cdot[(\mathbf{v}\times\mathbf{B})\,p] + (\gamma k_BT/m)\,\nabla_{\mathbf{v}}^2\,p.$$

Using the fact that $\nabla_{\mathbf{v}} \cdot (\mathbf{v} \times \mathbf{B}) \equiv 0$, this equation simplifies to yield

$$\partial p/\partial t = \gamma \nabla_{\mathbf{v}} \cdot (\mathbf{v} p) - (q/m)(\mathbf{v} \times \mathbf{B}) \cdot (\nabla_{\mathbf{v}}\, p) + (\gamma k_B T/m)\, \nabla_{\mathbf{v}}^2\, p.$$

35. We find, for $t \geq 0$,

$$\langle v_i(0)\, v_j(t) \rangle = \frac{k_B T}{m}\, e^{-\gamma t} \left[n_i\, n_j + (\delta_{ij} - n_i\, n_j)\, \cos\, \omega_{\mathrm{c}} t - \epsilon_{ijk}\, n_k\, \sin\, \omega_{\mathrm{c}} t \right].$$

For all t, we then have

$$\langle v_i(0)\, v_j(t) \rangle = \frac{k_B T}{m}\, e^{-\gamma |t|} \left[n_i\, n_j + (\delta_{ij} - n_i\, n_j)\, \cos\, \omega_{\mathrm{c}} t - \epsilon_{ijk}\, n_k\, \sin\, \omega_{\mathrm{c}} t \right].$$

36. Hence

$$\langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle = \frac{k_B T}{m}\, e^{-\gamma |t|}\, (1 + 2 \cos\, \omega_{\mathrm{c}} t),$$
$$\langle \mathbf{v}(0) \times \mathbf{v}(t) \rangle = -\frac{2 k_B T \mathbf{n}}{m}\, e^{-\gamma |t|}\, \sin\, \omega_{\mathrm{c}} t.$$

37. $D_{ij} = \dfrac{k_B T}{m\gamma} \left\{ n_i\, n_j + (\delta_{ij} - n_i\, n_j)\, \dfrac{\gamma^2}{\gamma^2 + \omega_{\mathrm{c}}^2} \right\}.$

38. When $\mathbf{n} = (0, 0, 1)$, the off-diagonal elements of D_{ij} vanish, while the diagonal elements are given by

$$D_{33} = D_{\mathrm{long}} = \frac{k_B T}{m\gamma}\,, \quad D_{11} = D_{22} = D_{\mathrm{trans}} = \frac{k_B T}{m\gamma} \left(\frac{\gamma^2}{\gamma^2 + \omega_{\mathrm{c}}^2} \right).$$

The diffusion coefficient in the direction of the field is therefore unaffected by the field, while that in the plane transverse to the field is attenuated by the factor $\gamma^2/(\gamma^2 + \omega_{\mathrm{c}}^2)$. This is a consequence of the tendency of the magnetic field to turn the transverse component of the velocity around, leading to orbital or cyclotron motion in the absence of thermal fluctuations.

Since $\langle x^2(t) \rangle \sim 2 D_{11}\, t$, $\langle y^2(t) \rangle \sim 2 D_{22}\, t$, $\langle z^2(t) \rangle \sim 2 D_{33}\, t$, we get

$$\langle r^2(t) \rangle \sim 6 D_{\mathrm{eff}}\, t, \quad \text{where} \quad D_{\mathrm{eff}} = \frac{k_B T}{m\gamma} \left(\frac{\gamma^2 + \frac{1}{3}\omega_{\mathrm{c}}^2}{\gamma^2 + \omega_{\mathrm{c}}^2} \right).$$

39. The dynamic mobility tensor in the presence of a magnetic field is given by

$$\mu_{ij}(\omega) = \frac{1}{m}\Big\{\frac{n_i\,n_j}{(\gamma - i\omega)} + \frac{(\delta_{ij} - n_i\,n_j)(\gamma - i\omega)}{(\gamma - i\omega)^2 + \omega_{\rm c}^2} - \frac{\epsilon_{ijk}\,n_k\,\omega_{\rm c}}{(\gamma - i\omega)^2 + \omega_{\rm c}^2}\Big\}.$$

The first two terms in the curly brackets represent the symmetric part of the tensor, while the third term represents the antisymmetric part.

40. $$\frac{\partial \rho}{\partial t} = -D\,n(n-1)\,\frac{\partial}{\partial \xi}\,[\xi^{(n-2)/n}\,\rho] + D\,n^2\,\frac{\partial^2}{\partial \xi^2}\,[\xi^{2(n-1)/n}\,\rho].$$

41. $$d\xi(t) = D\,n(n-1)\,\xi^{(n-2)/n}\,dt + n\sqrt{2D}\;\xi^{(n-1)/n}\,dW(t).$$

42. Set $t = 0$ in the differential equations satisfied by $p_{\rm R}(x,t)$ and $p_{\rm L}(x,t)$ and use the initial conditions $p_{\rm R}(x,0) = \delta(x),\ p_{\rm L}(x,0) = 0$. This gives

$$\big[\partial p_{\rm R}(x,t)/\partial t\big]_{t=0} = -c\,\delta\,'(x) - \nu\,\delta(x), \quad \big[\partial p_{\rm L}(x,t)/\partial t\big]_{t=0} = \nu\,\delta(x),$$

where $\delta\,'(x) = (d/dx)\,\delta(x)$. Hence the initial conditions on the positional PDF $p(x,t)$ are

$$p(x,0) = \delta(x),\ \big[\partial p(x,t)/\partial t\big]_{t=0} = -c\,\delta\,'(x).$$

43. The first moment $\langle X(t)\rangle$ satisfies the differential equation

$$\Big(\frac{d^2}{dt^2} + 2\nu\,\frac{d}{dt}\Big)\langle X(t)\rangle = 0.$$

The second moment $\langle X^2(t)\rangle$ satisfies the differential equation

$$\Big(\frac{d^2}{dt^2} + 2\nu\,\frac{d}{dt}\Big)\langle X^2(t)\rangle = 2c^2.$$

44. The initial conditions on the first and second moments of $X(t)$ are

$$\langle X(0)\rangle = 0,\ \big[d\langle X(t)\rangle/dt\big]_{t=0} = c$$

and

$$\langle X^2(0)\rangle = 0,\ \big[d\langle X^2(t)\rangle/dt\big]_{t=0} = 0.$$

Hence the solutions for $\langle X(t)\rangle$ and $\langle X^2(t)\rangle$ are

$$\langle X(t)\rangle = \frac{c}{2\nu}\,\big(1 - e^{-2\nu t}\big)$$

and

$$\langle X^2(t)\rangle = \frac{c^2}{2\nu^2}\left(2\nu t - 1 + e^{-2\nu t}\right).$$

The variance of $X(t)$ is therefore

$$\operatorname{Var} X(t) = \langle X^2(t)\rangle - \langle X(t)\rangle^2 = \frac{c^2}{4\nu^2}\left(4\nu t - 3 + 4e^{-2\nu t} - e^{-4\nu t}\right).$$

Note that the long-time behaviour of the variance is given by $\operatorname{Var} X(t) \sim (c^2/\nu)t$, which is linear in t. The process is therefore diffusive. With the identification $D = \lim c^2/(2\nu)$ in the limit $c \to \infty,\ \nu \to \infty$, this is in complete agreement with the familiar result $\operatorname{Var} X(t) \sim 2Dt$ for ordinary diffusion in one spatial immersion.

45. $I_j(u) \sim e^u/(2\pi u)^{1/2}$ as $u \to \infty$ for all finite $j \quad \Rightarrow \langle N(0;0,T)\rangle \sim (2\nu T/\pi)^{1/2}$.

46. $\lim_{T\to\infty} [\Delta N(0;0,T)/\langle N(0;0,T)\rangle] = (\frac{1}{2}\pi - 1)^{1/2}$.

47. The recursion relations satisfied by the mean first passage times $\{T_k\}$ are

$$\begin{aligned} T_k &= \tfrac{1}{2}\left(T_{k-1} + T_{k+1}\right) + 1 \quad (1 \le k \le j-2) \\ T_0 &= T_1 + 1 \\ T_{j-1} &= \tfrac{1}{2}\,T_{j-2} + 1. \end{aligned}$$

The solution is $T_k = j^2 - k^2,\ \ 0 \le k \le j$.

48. The generating function of $P(0, n \,|\, 0)$ is

$$\pi_{00}(z) = \frac{1 - \sqrt{1-z^2}}{\sqrt{1-z^2}}\,.$$

The generating function of $F(0, n \,|\, 0)$ is

$$\phi_{00}(z) = \frac{\pi_{00}(z)}{1 + \pi_{00}(z)} = 1 - \sqrt{1-z^2}.$$

Since $\phi_{00}(1) = 1$, a return to the origin is a sure event (i.e., it occurs with probability 1). This is true for a return to any other site as well, by the translational invariance of the infinite lattice. Hence an unbiased random walk on a linear lattice is *recurrent*.

The mean recurrence time, however, is infinite because

$$\langle t(0\,|\,0)\rangle = -\Big[\frac{d\phi_{00}(z)}{dz}\Big]_{z=1} = \infty.$$

Hence the random walk is *null-recurrent*.

49. The generating function of $P(1,n\,|\,0)$ is

$$\pi_{10}(z) = \frac{1-\sqrt{1-z^2}}{z\sqrt{1-z^2}}\,.$$

The generating function of $F(1,n\,|\,0)$ is

$$\phi_{10}(z) = \frac{\pi_{10}(z)}{1+\pi_{11}(z)} = \frac{1-\sqrt{1-z^2}}{z}$$

using the fact that $\pi_{11}(z) = \pi_{00}(z)$.

Once again, the mean first-passage time from 0 to 1 is infinite, because

$$\langle t(1\,|\,0)\rangle = -\Big[\frac{d\phi_{10}(z)}{dz}\Big]_{z=1} = \infty.$$

50. The recurrence-time distribution is given by

$$F(0,2n\,|\,0) = \frac{\Gamma\big(n-\frac{1}{2}\big)}{2\sqrt{\pi}\,n!}\,, \quad F(0,2n-1\,|\,0) = 0, \quad \text{where } \; n \geq 1.$$

The first-passage-time distribution is given by

$$F(1,2n-1\,|\,0) = \frac{\Gamma\big(n-\frac{1}{2}\big)}{2\sqrt{\pi}\,n!}\,, \quad F(1,2n\,|\,0) = 0, \quad \text{where } \; n \geq 1.$$

Simplifying the gamma function, we may also write

$$F(0,2n\,|\,0) = F(1,2n-1\,|\,0) = \frac{(2n-2)!}{2^{2n-1}\,(n-1)!\,n!}\,, \quad n \geq 1.$$

Observe that $F(0,2n\,|\,0)$ happens to be equal to $F(1,2n-1\,|\,0)$, although the respective random walk paths contributing to the two distributions are quite distinct from each other.

51. Multiply both sides of the equation $\Delta_{kl}\, F(n,l) = F(n+1,k) - F(n,k)$ by n, and sum over n. Then

$$\begin{aligned}\Delta_{kl} T_l^{(1)} &= \sum_{n=0}^{\infty} nF(n+1,k) - T_k^{(1)} \\ &= \sum_{n=0}^{\infty} (n+1-1)F(n+1,k) - T_k^{(1)} = T_k^{(1)} - 1 - T_k^{(1)} = -1.\end{aligned}$$

52. Similarly, multiplying both sides of the equation by n^q and summing over n gives

$$\Delta_{kl} T_l^{(q)} = \sum_{n=0}^{\infty} n^q F(n+1,k) - T_k^{(q)}.$$

But $n^q = (n+1-1)^q = \sum_{r=0}^{q} \binom{q}{r} (-1)^r (n+1)^{q-r}$. Hence

$$\Delta_{kl} T_l^{(q)} = \sum_{r=1}^{q} \binom{q}{r} (-1)^r\, T_k^{(q-r)}, \quad q \geq 1.$$

Hence $\Delta_{kl} T_l^{(q)}$ is equal to a linear combination of the lower moments $T_k^{(r)}$ where $r = 0, 1, 2, \ldots, (q-1)$. In particular, the second moment of the time-to-trapping satisfies the equation

$$\Delta_{kl} T_l^{(2)} = -2T_k^{(1)} + 1.$$

53. In this case $\nu_k = \frac{1}{2}$ (each site has two nearest-neighbor sites), and the backward Kolmogorov equation for the mean first-passage time is

$$\Delta_{kl}\, T_l^{(l)} = \tfrac{1}{2}\big(T_{k-1}^{(1)} + T_{k+1}^{(1)}\big) - T_k^{(1)} = -1.$$

Hence the second difference of $T_k^{(1)}$ is a constant (i.e., it is independent of k), which implies that $T_k^{(1)}$ can only be a quadratic function of k, at best. Moreover, $T_k^{(1)} = T_{-k}^{(1)}$, by an obvious symmetry (the traps are at $\pm j$). Therefore $T_k^{(1)}$ must be of the form $ak^2 + b$. Since $T_j^{(1)} \equiv 0$, we have $b = -aj^2$, so that $T_k^{(1)} = a(k^2 - j^2)$. Setting $j = 1$, the only non-trap site is $k = 0$, and it is obvious that $T_0^{(1)} = 1$ in this case. This gives $a = -1$. It follows that the general solution is

$$T_k^{(1)} = j^2 - k^2.$$

Note, in particular, that $T_0^{(1)} = j^2$. In other words, the mean time to reach a site at a *distance* j from the origin for the first time is just j^2, as expected.

54. The equation $\Delta_{kl} T_l^{(2)} = -2T_k^{(1)} + 1$ becomes, in the present instance,

$$\tfrac{1}{2}\big(T_{k-1}^{(2)} + T_{k+1}^{(2)}\big) - T_k^{(2)} = 2(k^2 - j^2) + 1.$$

As the second difference of $T_k^{(2)}$ is quadratic in k, $T_k^{(2)}$ itself must be a quartic in k. Since it must be a symmetric function of k, we have the general form $T_k^{(2)} = ak^4 + bk^2 + c$. The boundary condition $T_k^{(2)} = 0$ leads to $T_k^{(2)} = a(k^4 - j^4) + b(k^2 - j^2)$. Substituting this in the difference equation satisfied by $T_k^{(2)}$, we get $a = \frac{1}{3}$ and $b = \frac{2}{3} - 2j^2$. The solution for $T_k^{(2)}$ is then

$$T_k^{(2)} = \tfrac{5}{3}j^4 - 2j^2k^2 + \tfrac{1}{3}k^4 + \tfrac{2}{3}(k^2 - j^2).$$

55. The variance of the time-to-trapping is therefore

$$\mathrm{Var}\,(t_k) = T_k^{(2)} - [T_k^{(1)}]^2 = \tfrac{2}{3}(j^2 - k^2)(j^2 + k^2 - 1).$$

The relative fluctuation in t_k is then

$$\frac{\Delta t_k}{\langle t_k \rangle} = \Big[\frac{2}{3}\Big(\frac{j^2 + k^2 - 1}{j^2 - k^2}\Big)\Big]^{1/2}, \quad -(j-1) \le k \le (j-1).$$

Note that, for a random walk starting at the origin ($k = 0$), the relative fluctuation in the first-passage time is $\sqrt{(2/3)}$, independent of j. Likewise, if $j \to \infty$ while k remains finite, the relative fluctuation tends to $\sqrt{(2/3)}$.

56. $\widetilde{W}(n,s) = \dfrac{[1 - \widetilde{\psi}(s)]}{s}\,[\widetilde{\psi}(s)]^n.$

57. $\widetilde{L}(z,s) = \dfrac{[1 - \widetilde{\psi}(s)]}{s\,[1 - g(z)\,\widetilde{\psi}(s)]}$, where $g(z) = \frac{1}{2}\,(z + z^{-1})$.

58. $\mathcal{L}[\langle j(t)\rangle] = \big[\partial \widetilde{L}(z,s)/\partial z\big]_{z=1} = 0$, since $\big[dg(z)/dz\big]_{z=1} = 0$. Hence the mean displacement $\langle j(t)\rangle = 0$ for all t, as it must for an unbiased random walk.

Now, $\big[\partial^2 \widetilde{L}(z,s)/\partial z^2\big]_{z=1} = \mathcal{L}[\langle j(j-1)(t)\rangle] = \mathcal{L}[\langle j^2(t)\rangle]$, since $\langle j(t)\rangle = 0$. We then find

$$\mathcal{L}[\langle j^2(t)\rangle] = \frac{\widetilde{\psi}(s)}{s\,[1 - \widetilde{\psi}(s)]}\,.$$

This is an exact result.

59. The $t \to \infty$ behavior of $\langle j^2(t)\rangle$ is determined by the $s \to 0$ behavior of $\mathcal{L}[\langle j^2(t)\rangle]$. Using the fact that

$$\widetilde{\psi}(s) = 1 - s\tau + \mathcal{O}(s^2)$$

in the neighbourhood of $s = 0$, we get at once

$$\mathcal{L}[\langle j^2(t)\rangle] \sim \frac{1}{s^2\,\tau}\,, \;\; \text{so that} \;\; \langle j^2(t)\rangle \sim \frac{t}{\tau}\,.$$

In other words, a CTRW whose waiting-time density $\psi(t)$ has a finite first moment τ leads to normal diffusive behavior at long times.

60. If, near $s = 0$, we have $\widetilde{\psi}(s) = 1 - cs^\gamma +$ higher powers of s (where $0 < \gamma < 1$), it follows that the leading term in $\mathcal{L}[\langle j^2(t)\rangle]$ is $1/(c\,s^{\gamma+1})$. Hence the inverse Laplace transform has a leading long-time behavior given by

$$\langle j^2(t)\rangle \sim t^\gamma,$$

i.e., it is *sub-diffusive*. This is an instance of **anomalous diffusion**. Note that the small-s behavior of $\widetilde{\psi}(s)$ given above arises when $\psi(t)$ has a power-law tail, i.e., it decays to zero like $1/t^{1+\gamma}$ as $t \to \infty$.

61. The sojourn probability in L is

$$\begin{aligned}\widetilde{w}_n =& P\big(L, n-1; L, n-2; \cdots; L, 1; L, 0\big) \\ =& \int_L dx_0 \cdots \int_L dx_{n-2} \int_L dx_{n-1}\, \rho(x_0)\, \delta\big(x_1 - f(x_0)\big) \cdots \delta\big(x_{n-1} - f(x_{n-2})\big).\end{aligned}$$

Since $f(x) = 2x$ when $x \in L$, and $\rho(x_0) = 1$, this reduces to

$$\widetilde{w}_n = \int_0^{\frac{1}{2}} dx_0 \cdots \int_0^{\frac{1}{2}} dx_{n-2} \int_0^{\frac{1}{2}} dx_{n-1}\, \delta(x_1 - 2x_0) \cdots \delta(x_{n-1} - 2x_{n-2}).$$

62. When each integral in the foregoing (starting with the integration over x_{n-1} is evaluated using a δ-function, the upper limit of integration in the next integral is shrunk by a factor of $\frac{1}{2}$. The final result is

$$\widetilde{w}_n = \int_0^{1/2^n} dx_0 = \frac{1}{2^n}\,.$$

Hence the normalized probability distribution of sojourn in L is

$$\widetilde{H}_n = \frac{\widetilde{w}_{n+1}}{\widetilde{w}_1} = \frac{\widetilde{w}_{n+1}}{\mu_L} = \frac{1}{2^{n+1}} \times \frac{2}{1} = \frac{1}{2^n}\,.$$

Since $\sum_1^\infty \widetilde{H}_n = \sum_1^\infty 2^{-n} = 1$, $\widetilde{H}_n$ is properly normalized, as required.

63. The distribution of the time of escape out of L is given by

$$\widetilde{E}_n \stackrel{\text{def.}}{=} P(R, n; L, n-1; \cdots; L, 1|L, 0) = \widetilde{H}_{n-1} - \widetilde{H}_n = \frac{1}{2^{n-1}} - \frac{1}{2^n} = \frac{1}{2^n}\,.$$

Again, the normalization condition $\sum_1^\infty \widetilde{E}_n = 1$ is satisfied. Hence the mean time of escape out of L is $\sum_{n=1}^\infty n/2^n = 2$.

64. In terms of the quantities $\{\widetilde{w}_n\}$, the distribution of the time of recurrence to R works out to

$$R_n \stackrel{\text{def.}}{=} P(R, n; L, n-1; \cdots; L, 1|R, 0) = \frac{(\widetilde{w}_{n-1} - 2\widetilde{w}_n + \widetilde{w}_{n+1})}{(\widetilde{w}_0 - \widetilde{w}_1)}\,,$$

where $\widetilde{w}_0 \equiv 1$ and $\widetilde{w}_1 = \mu_L$. Therefore the denominator in the expression above is $\widetilde{w}_0 - \widetilde{w}_1 = \mu_R$. Since $\widetilde{w}_n = 1/2^n$, we get $R_n = 1/2^n$ in this instance. Once again, it is obvious that the normalization condition $\sum_1^\infty R_n = 1$ is satisfied. Further, the mean time of recurrence to R is given by $\sum_{n=1}^\infty n/2^n = 2 = 1/\mu_R$, in accordance with Poincaré's recurrence theorem.

65. The map $f(x) = 2x \mod 1$ comprises two branches, namely, $f_L(x) = 2x$ when $x \in L$ and $f_R = 2x - 1$ when $x \in R$.

The one-step conditional probabilities are as follows:

$$\begin{aligned} P(L, 1|L, 0) &= \frac{P(L, 1; L, 0)}{P(L)} = \frac{1}{\mu_L} \int_0^{\frac{1}{2}} dx_0 \int_0^{\frac{1}{2}} dx_1\, \rho(x_0)\, \delta\big(x_1 - f_L(x_0)\big) \\ &= 2 \int_0^{\frac{1}{4}} dx_0 = \tfrac{1}{2}\,, \end{aligned}$$

$$\begin{aligned} P(L, 1|R, 0) &= \frac{P(L, 1; R, 0)}{P(R)} = \frac{1}{\mu_R} \int_{\frac{1}{2}}^1 dx_0 \int_0^{\frac{1}{2}} dx_1\, \rho(x_0)\, \delta\big(x_1 - f_R(x_0)\big) \\ &= 2 \int_{\frac{1}{2}}^{\frac{3}{4}} dx_0 = \tfrac{1}{2}\,, \end{aligned}$$

$$
\begin{aligned}
P(R,1|L,0) &= \frac{P(R,1;L,0)}{P(L)} = \frac{1}{\mu_L}\int_0^{\frac{1}{2}} dx_0 \int_{\frac{1}{2}}^1 dx_1\, \rho(x_0)\,\delta\big(x_1 - f_L(x_0)\big) \\
&= 2\int_{\frac{1}{4}}^{\frac{1}{2}} dx_0 = \tfrac{1}{2}\,, \\
P(R,1|R,0) &= \frac{P(R,1;R,0)}{P(R)} = \frac{1}{\mu_R}\int_{\frac{1}{2}}^1 dx_0 \int_{\frac{1}{2}}^1 dx_1\, \rho(x_0)\,\delta\big(x_1 - f_R(x_0)\big) \\
&= 2\int_{\frac{3}{4}}^{1} dx_0 = \tfrac{1}{2}\,.
\end{aligned}
$$

66. Consider the stationary Markov chain generated by the transition matrix

$$
W = \begin{pmatrix} P(L,1|L,0) & P(L,1|R,0) \\ P(R,1|L,0) & P(R,1|R,0) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.
$$

Observe that W is idempotent, i.e., $W^n = W$ for any positive integer n. Hence, for each $n \geq 1$ the conditional probabilities are given by

$$
\begin{pmatrix} P(L,n|L,0) & P(L,n|R,0) \\ P(R,n|L,0) & P(R,n|R,0) \end{pmatrix} = W^n = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.
$$

Now consider the distribution of the time of recurrence to R. It is defined as

$$
R_n \stackrel{\text{def.}}{=} P(R,n;L,n-1;\cdots;L,1|R,0) = \frac{P(R,n;L,n-1;\cdots;L,1;R,0)}{P(R)}\,.
$$

But the joint probability $P(R,n;L,n-1;\cdots;L,1;R,0)$ is identically equal to

$$
P(R,n|L,n-1;\cdots;L,1;R,0)P(L,n-1;\cdots;L,1;R,0).
$$

Owing to the Markov property this product simplifies to

$$
P(R,n|L,n-1)P(L,n-1;\cdots;L,1;R,0) = \tfrac{1}{2}P(L,n-1;\cdots;L,1;R,0).
$$

Reducing the joint probability $P(L,n-1;\cdots;L,1;R,0)$ in a similar manner, and repeating the process, we finally arrive at

$$
\begin{aligned}
P(R,n;L,n-1;\cdots;L,1;R,0) &= \frac{1}{2^{n-1}}P(L,1;R,0) \\
&= \frac{1}{2^{n-1}}P(L,1|R,0)P(R) = \frac{1}{2^n}P(R).
\end{aligned}
$$

Therefore $R_n = 1/2^n$. But this is precisely the result already obtained by a direct evaluation of the multiple integral resulting from the definition of R_n. This corroborates the claim that the partitioning of the unit interval into the cells L and R makes the coarse-grained dynamics of the Bernoulli map equivalent to a Markov chain governed by the transition matrix W.