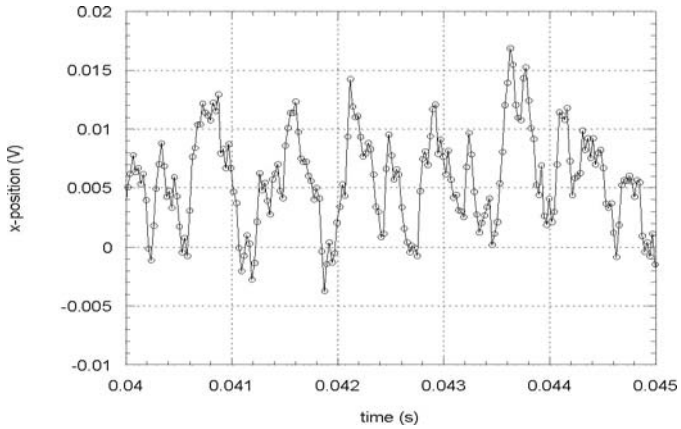
A Minimum of Stochastics for Scientists

Ten Lectures



Noel Corngold

Caltech
2009

2nd corrected printing
© 2009 Noel Corngold

Contents

Preface

These pages contain material I have tried to convey, in a course given occasionally, to a Caltech audience composed mostly of graduate students. The idea behind the course was to introduce students to the ideas and attitudes that underlie the statistical modeling of physical, chemical, biological systems. To provide a sufficient minimum to begin the repair of total ignorance, to get one started, to avoid embarrassing moments at an oral qualifying examination, perhaps. This little book might be titled "An Introduction to an Introduction to . . ." The student who wishes to go deeper will find much that is rich in the classic texts and essays listed under References.

If the presentation here is clear and helpful, it will be in large part due to the efforts of Emily Adelsohn, to whom I am most grateful.

Noel Corngold Pasadena 2009

Lecture 1

General Considerations

When a leaf drops, its trajectory is hardly that of the simple point-mass beloved of homework exercises. Aside from issues regarding shape, we do not know, cannot know, the complicated, aerodynamic forces acting upon the leaf. And if the experiment is repeated, the repetitions will yield a variety of outcomes. In effect, the motion is incalculable. (The issue is, of course, imperfect knowledge.) What to do? How do we systematize, organize, the results of repeated experiments? Is prediction possible? Enter probability.

Probability concerns the possible outcomes of a venture, an experiment. Possible outcomes? In the macroscopic world we have Newton's equations of motion, we have Maxwell's equations of electro-magnetism. Given the required initial information and the correct boundary conditions, we expect a unique outcome. Why do we need probability? We will find it to be a powerful tool in systematizing uncertainty.(Probabilistic issues are indeed central to quantum mechanics—but that is another story.)

We quantify the situation by introducing an appropriate probability-function or distribution. The quantity whose outcome is in question is often called a stochastic variable (*stochastic* = random, pertaining to conjecture). We will denote such variables by capital letters, their possible values by lower case letters. Thus, the variable *X*. Its *N* possible outcomes are $x_1, x_2, \ldots x_j, \ldots x_N$ We associate a *probability* $p(x_j) \geq 0$ with each outcome. The probabilities are real, non-negative numbers and are chosen ("normalized") so that

$$\sum_j p(x_j) = 1 . \qquad (1.1)$$

The probabilities sum to unity. When the outcomes may be regarded as varying continuously over some interval, say $a \leq x \leq b$, we ask for the probability that the outcome lies in the interval *(x, x+dx)*. That quantity is written *p(x) dx,* and we have introduced the *probability density p(x).* Its normalization is:

$$\int_a^b dx\, p(x) = 1 . \qquad (1.2)$$

Although the discrete and continuous functions differ in their physical dimensions, we will not distinguish between them, notationally. The differences should be apparent from the context of the discussion. And, occasionally, we will use the notation *Pr{X=x)* for *p(x).*

While the generalization of these ideas to more than one variable is obvious—e.g., what is the probability of finding a diffusing atom in the element *dx dy dz* with spin

"up?"—several new ideas appear. Suppose we have two stochastic variables, (X_1, X_2) with their associated distribution $p(x_1, x_2)$. We would read the expression as "the probability density for X_1 having the value x_1 and X_2 the value x_2."

The first idea is that of the *reduced, or marginal distribution*

$$p_1(x_1) = \int dx_2 \; p(x_1, x_2) \qquad p_2(x_2) = \int dx_1 \; p(x_1, x_2) \,. \tag{1.3}$$

Each is properly normalized. (We have "integrated or summed over" an unimportant variable.) The second idea is that of statistical independence. We call our two variables (X_1, X_2) *independent* if $p(x_1, x_2) = p_1(x_1)\, p_2(x_2)$. In the general case of *N* variables, should the set $X_1, X_2 \ldots X_m$ be independent of the others, we would have

$$p(x_1, x_2, \ldots x_N) = p_m(x_1, x_2, \ldots x_m)\, p_{N-m}(x_{m+1}, x_{m+2}, \ldots x_N) \tag{1.4}$$

Since the last example deals with a large number of stochastic quantities, indexed by a discrete subscript, we might mention that we will later encounter an infinite number of variables indexed by a continuous variable. Thus $p(x_1, x_2, \ldots x_N) \to p(x(s)), \quad 0 \le s \le 1$ and there is a stochastic variable *X(s)* for each value of s in the interval. Application might be to stochastic "fields" —perhaps to a treatment of turbulence in fluids.

When we treat dependence and independence, the notion of *conditional probability* is useful. In the case of two variables, we write (*Bayes' Rule)*

$$p(x_1, x_2) = p_{12}(x_1 \mid x_2)\, p_2(x_2) = p_{21}(x_2 \mid x_1)\, p_1(x_1) \,. \tag{1.5}$$

The quantities $p(x_i \mid x_j)$ express the probability that X_i will yield the outcome x_i given that X_j has the value x_j. One might say that $p(x_1, x_2)$ describes an "and" situation, $p(x_1 \mid x_2)$ an "if." Should the variables be independent, we will have $p_{12}(x_1 \mid x_2) = p_1(x_1)$ and $p_{21}(x_2 \mid x_1) = p_2(x_2)$, with obvious meaning. The normalization of the conditional probabilities,

$$\int dx_1 \; p_{12}(x_1 \mid x_2) = 1 \qquad \int dx_2 \; p_{21}(x_2 \mid x_1) = 1 \tag{1.6}$$

has clear meaning, too. And the idea of conditional probability can be extended to more than two variables. It is particularly useful in the physics of many-body systems, where it is natural to ask about the probability of finding a particle at some point *P'* in the neighborhood of a particle known to be at point *P*. We may occasionally omit the subscripts in expressions like Eq. (1.5), when they are not required for clarity.

Finally, it frequently occurs that we are interested in the probability distribution associated with a stochastic variable *F* which is itself a function of several X_j and which has outcomes *f*. Thus, we seek *P(f)df.* That quantity is

$$P(f)=\int dx_1 ...dx_N \; p(x_1,...x_N)\; \delta(f - F(x_1,...x_N)) . \qquad (1.7)$$

A simple version of this expression, when only a single variable is involved is

$$P(f) \mid \frac{dF(x)}{dx} \mid_{x_0} = p(x_0) , \qquad (1.8)$$

with x_0 determined by $F(x_0) = f$. (We assume that the latter equation has but one solution.)

Origins and Examples

How do we know which probability distribution to use in a given situation? One obvious and a priori source might be an argument based on symmetry, or complete ignorance. The latter notion is often called the Principle of Insufficient Reason. For example, the probability of choosing a particular card from a randomized deck is 1/52; the probability that a needle, spun on a perfect mount, will find its point in the interval $(\theta, \theta + d\theta)$ is $\frac{1}{2\pi} d\theta$. (Perhaps these are merely definitions of "randomized deck," and "perfect mount.") Another example: if messages (or particles) arrive "in a completely random manner" during an extended interval *T*, the probability of arrival in the interval *(t, t + dt)* is $\frac{1}{T} dt$. Although these examples appear trivial, we shall be using them, combined with the notion of statistical independence, to derive quite important distributions. Then the issue of counting, the subject of *combinatorics*, will appear.

A more profound or fundamental definition of probability is based on the underlying experiment itself. We imagine the experiment carried out *N* times. Then, the probability assigned to the outcome "*j*" is

$$p(j)= Lim_{N\to\infty} \frac{number\ of\ trials\ yielding\ "j"}{N} , \qquad (1.9)$$

the quantity on the rhs being called, alternatively, the *frequency* of *j*. This is the definition we keep in mind in difficult times. And the possible outcomes need not be as simple as which side of a die will turn up. One might ask for the probability that a certain condition is fulfilled in the experiment. The outcomes then are merely two, success or failure, but the calculation of the probability might be challenging. The definition also leads us to conclude that if *j* and *k* are independent outcomes then the probability of either *j* or *k* or both is *p(j) + p(k).*

An interesting way to construct an appropriate probability distribution is to maximize the entropy associated with a given distribution, subject to constraints which reflect prior knowledge. (For further discussion see Appendix A.)

Features of the Probability Distribution

Most distributions are characterized by only a few features. We discuss, for the most part, stochastic variables having a continuous range of outcomes; the discrete case is quite similar. Particularly important are the *moments,* denoted μ'_n in the case of a single variable.

$$\mu'_n = \int dx\, x^n\ p(x) = < X^n > \quad \text{or} \quad \mu'_n = \sum_{j=1}^{N} (x_j)^n\, p(x_j) \qquad n = 0, 1, 2, \ldots \tag{1.10}$$

(Note that the *n*th moment of X may also be regarded as the average of the stochastic variable X^n.) Often, a few moments can be extracted from experiment. To what extent knowledge of the moments, positive and negative, uniquely define the underlying distribution function is a question that has intrigued mathematicians as well as scientists. In any case the zeroth moment is unity, and the first moment, the average of X, is often called the *mean* or *expectation*. One gets a clearer view of the distribution by describing higher moments relative to the mean. These are the un-primed μ_n,

$$\mu_n = \int dx\, (x - < x >)^n\ p(x)\ . \tag{1.11}$$

Thus, $\mu_0 = 1,\ \mu_1 = 0.$ The second moment

$$\mu_2 \equiv \sigma^2\ = \int dx\, (x - < x >)^2\ p(x) = \mu'_2 - \mu'^2_1\ , \tag{1.12}$$

called the *variance* or *dispersion*, gives insight into the spread of the distribution about its average value. (The square root of the variance is sometimes called the *half-width*.) The asymmetry, or "skewness," of a distribution may be characterized by

$$\gamma_1 = \frac{\mu_3}{\sigma^3}\ . \tag{1.13}$$

The higher moments, by their nature, give information about the farther reaches, the "tails" of distributions. We shall often deal with distributions having zero mean; they and their moments will be referred to as *centered.*

Many interesting distributions have only a few finite moments. A notable example is the Cauchy distribution (Lorentzian line-shape), $p(x) = \frac{1}{\pi}\frac{1}{1+x^2} \qquad -\infty < x < \infty$, encountered in a variety of spectroscopies. It is normalized and symmetric, but no positive moments exist. The famous Gaussian, or normal distribution, is a healthier example. Relative to its average (i.e., $<X> = 0$) it is

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp[-\frac{x^2}{2\sigma^2}] \qquad -\infty < x < \infty \tag{1.14}$$

and is characterized by its second moment. Its odd moments vanish, by symmetry, and its even moments are

$$\mu_n = < X^{2n} > = \frac{1}{2^n}\frac{(2n)!}{n!}\sigma^{2n}. \tag{1.15}$$

The Gaussian, as $\sigma \to 0$, provides a representation of Dirac's *delta-function* $\delta(x)$.

A useful parameter, the *kurtosis*, measures the deviation of the fourth moment of an arbitrary distribution from the simple value associated with the Gaussian. It is

$$\gamma_2 = \frac{\mu_4}{(\mu_2)^2} - 3. \tag{1.16}$$

Characteristic functions are very helpful in the calculation of moments. Consider the Fourier transform of the distribution,

$$C(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx}\, p(x) = < e^{-ikX} >. \tag{1.17}$$

Its power series expansion is

$$C(k) = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} < X^n >. \tag{1.18}$$

Thus, if we have the Fourier transform, we can read the moments. The classic example is the centered Gaussian considered above, where

$$C(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx} \quad \frac{1}{\sqrt{2\pi}\,\sigma}\exp[-\frac{x^2}{2\sigma^2}] \;=\; e^{-\frac{1}{2}k^2\sigma^2} \tag{1.19}$$

and the results quoted above emerge easily. Eq (1.19) will be used often during these lectures.) Of course one is not limited to Fourier transforms. Several other integral representations would work as well. A variant of the scheme, appropriate to discrete outcomes, introduces *generating functions.* Let the outcomes be non-negative integers, n with associated probabilities *p(n)*. It is often possible to find a function, *G(z)*, whose Taylor expansion is

$$G(z)=\sum_{n=0}^{\infty} p(n)\, z^n \ . \tag{1.20}$$

Then, the moments appear via $\mu'_0 = 1 = G(1), \quad \mu'_1 = G'(1), \quad \mu'_2 = G''(1) + G'(1),$ the primes referring to derivatives.

Another set of moments, the *cumulants*, are often useful. These are defined by a simple relation between characteristic functions. Along with

$$A(k)=\sum_{n=0}^{\infty} \frac{(-ik)^n}{n!}\mu_n = < e^{-ikX} > \tag{1.21}$$

write

$$B(k)=\log A(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!}\chi_n . \tag{1.22}$$

Eq. (1.22) defines the cumulant moments χ_n. One compares series expansions to get them. In terms of centered moments, we find,

$$\chi_1 = \mu_1 = 0, \quad \chi_2 = \mu_2 = \sigma^2 , \quad \chi_3 = \mu_3, \quad \chi_4 = \mu_4 - 3\mu_2^2 , \quad \ldots \tag{1.23}$$

and the striking result that all cumulant moments of the Gaussian vanish, except for χ_2. And since the second cumulant is the dispersion, a characteristic function lacking a quadratic term, e.g., C*(k)=Exp[* $-(k^4 + ..)$ *]* , would not be acceptable.

Cumulants play an important, somewhat unexpected role in statistical mechanics. Recall that the partition function, which is the key to statistical thermodynamics, may be regarded as an average of $e^{-\beta H}$, where $\beta = (k_B T)^{-1}$ and H is the Hamiltonian. In fact, the Helmholtz free energy is given by $F = -\ k_B T\ \log < e^{-\beta H} >$. But if we treat H as a stochastic variable X and replace "k" with $-i\beta$ in Eq. (1.22), we find

$$-\beta F = \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!}\chi_n(H) . \tag{1.24}$$

Thus, if we perform a high-temperature (small β) expansion of the free energy, we find that the co-efficients of powers of inverse temperature are precisely the cumulant moments of the Hamiltonian. In studies of complex physical systems one finds often that cumulants will sum troubling terms into tidy packages.

Correlation Functions

The introduction and discussion of moments may be extended easily to the case of several stochastic variables. The most interesting objects encountered here are *correlation functions*. As an example, let (X_1, X_2) be stochastic variables each having mean value zero. Then

$$< X_1 X_2 >= \int dx_1 \int dx_2 \; p(x_1, x_2) \; x_1 x_2 \tag{1.25}$$

is a measure of the degree of correlation of the variables. Suppose that we are tracing the motion of a tagged particle as it diffuses through a fluid. We may denote its initial velocity by X_1, the velocity at a later time by X_2. Both will be treated as stochastic variables. The analysis becomes a bit more "physical" if we introduce the conditional probability. Thus

$$< X_1 X_2 >= \int dx_2 \int dx_1 \; p(x_2 | x_1) \; p_1(x_1) \; x_1 x_2 . \tag{1.26}$$

$p_1(x_1)$ will describe the distribution of initial velocities. If the system is at thermal equilibrium, the distribution will be Maxwellian—i.e., Gaussian, with zero mean. Should the particle initially have velocity x_1 the conditional probability gives the likelihood of having velocity x_2 some time later. (We remain, for simplicity, in one dimension.) The conditional probability is a sort of "propagator." As a function of x_2 it is "peaked" with regard to x_1 at short times, and "forgets," or is independent of x_1, at long times. We might model it as

$$p(x_2 | x_1) = \frac{1}{\sqrt{2\pi}\,\sigma(t)} \exp[-\frac{(x_2 - x_1 \phi(t))^2}{2\sigma(t)^2}] , \tag{1.27}$$

where the dispersion, $\sigma(t)$, grows continually from an initial value of zero, and $\phi(t)$, a relaxation function, describes the progressive amnesia. The latter diminishes continually from an initial value of unity, when $p(x_2 | x_1) = \delta(x_2 - x_1)$. With this model, the correlation integral can be evaluated easily by a change of variable $y = x_2 - x_1 \phi(t)$, to get

$$< X_1 X_2 >= < X_1^2 > \phi(t) . \tag{1.28}$$

If we use more familiar notation, and an exponential relaxation function, Eq. (1.28) becomes

$$< V V(t) >= < V^2 > e^{-\beta t} , \tag{1.29}$$

the averaged square velocity being the equilibrium value. It is proportional to temperature. Note that, somewhat surprisingly, the growing dispersion does not enter explicitly into the result.

We have used the terms "initially" and "time *t* later" in the previous discussion. We might have used times t_1 and t_2 instead, where $t_2 - t_1 = t$. These situations, in which only the difference in time arguments matters or, formally, where all features are invariant with respect to simultaneous translations of time variables, are called *stationary*. When the same feature appears in space variables, we shall describe the system as *homogeneous.*

Along with *cumulants* and *correlations* come *covariance* and *convolution.* The *covariance* of (X_1, X_2) is simply

$$< (X_1 - < X_1 >) \, (X_2 - < X_2 >) > = < X_1 \, X_2 > - < X_1 > < X_2 >, \qquad (1.30)$$

while the *convolution* describes something new. We are now concerned with the probability associated with the sum of independent stochastic variables. Suppose that the outcomes are discrete and positive, and let us begin with two variables. Thus $Y = X_1 + X_2$, and

$$p(y) = \sum_{x_1=0} \sum_{x_2=0} p_1(x_1) \, p_2(x_2) \, \delta(x_1 + x_2 - y) \qquad (1.31)$$

where (see Eq. (1.7)) we sum over the outcomes, and the "discrete delta-function" is either zero or unity. The generating function for *y* is

$$G(s) = \sum_y s^y \, p(y) = \sum_y \sum_{x1} \sum_{x2} p_1(x_1) \, p_2(x_2) \, s^y \, \delta(x_1 + x_2 - y), \qquad (1.32)$$

and we treat the rhs sum in two ways. First, we do the *y*-sum. The *x*-sums which result may be factored, to give $G(s) = g_1(s) \, g_2(s)$, the *g's* belonging to the corresponding *p's*. The interesting case $Y = X_1 + X_2 + ... X_N$, with the variables having identical statistics, leads then to $G(s) = g(s)^N$. An example: In binomial statistics, which we shall study soon, the *X's* refer to individual trials. Each trial has two outcomes, (*1,0*), graced by probabilities (*p,q*). Then *g(s) = (ps + q)* , and we have the familiar $G(s) = (p\,s + q)^N$.

In the second approach we sum on x_2 to get

$$G(s) = \sum_y s^y \, p(y) = \sum_y s^y \sum_{x=0}^{y} p_1(y - x) \, p_2(x) \quad , \qquad (1.33)$$

or,

$$p(y)=\sum_{x=0}^{y} p_1(y-x)\,p_2(x)=\sum_{x=0}^{y} p_1(x)\,p_2(y-x),$$

describing the *convolution* of two sequences, p_1 and p_2. The probability for y to occur is the sum of probabilities for mutually exclusive events, each describing a different "route to y."

In closing this lecture we should mention a situation where the array of probabilities we will be studying becomes particularly important. It occurs when we are measuring some property of a system, and the measurement, conducted over an interval of time, is in effect a time-average over some rapidly varying feature of the system. The measurement might extract details of fluctuations as well. At that point, the powerful theory of *ergodic* behavior comes into play. One may assert that under certain conditions the time average is equal to an ensemble average governed by a particular distribution function. The fluctuations also convey information about internal dynamics. Then, stochastics can enter and play a powerful role.

Lecture 2

The Binomial Distribution

We begin with a study of the simplest distribution, from which much will grow. The distribution is associated with the simple, random, coin-toss. The experiment consists of *N* repeated, independent trials of an exercise which has only two outcomes. The outcomes are described, variously, as "up-down" (e.g., a quantum-mechanical spin), "plus-minus," "success-failure." The 2^N possible outcomes of the *N* trials may be described by a discrete distribution with *N* two-valued variables, $p(\sigma_1, \sigma_2 ... \sigma_N)$, $\sigma_j = \pm$. Since the trials are independent we have

$$p(\sigma_1, \sigma_2 ... \sigma_N) = \prod_{m=1}^{N} p(\sigma_m). \tag{2.1}$$

The two probabilities assigned to each trial are denoted, traditionally, by *p* and *q*, $p + q = 1$. We might describe a typical outcome by a symbol with *N* entries (+ + - + - - +….), *k* "plus" and (*N-k)* "minus." Associated is the probability $p^k q^{N-k}$. Our primary interest is in the probability that there will be *k* successes in the *N* trials. These successes can occur in various sequences. These are independent outcomes and we add the probability associated with each. Since that probability is the same in each case, we need only count the number of ways the *k*-successes can occur. But that, clearly, is the number of ways one can arrange the *k* identical plus-signs among the *N* signs, namely

$$\frac{N!}{(N-k)!\,k!} = \binom{N}{k},$$

−the binomial co-efficient. The probability for *k* successes is, then

$$p(k, N) = \binom{N}{k} p^k q^{N-k} \tag{2.2}$$

and

$$\sum_{k=0}^{N} p(k, N) = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} = (p+q)^N = 1$$

as one would expect.

The generalization to three or more outcomes per trial is obvious. For example, should the associated probabilities be p_1, p_2, p_3 the probability, in *N* trials of n_1 "ones," n_2 "twos" and n_3 "threes," with $p_1 + p_2 + p_3 = 1$, $n_1 + n_2 + n_3 = N$ is

$$P(n_1, n_2, n_3; N) = \frac{N!}{n_1!\, n_2!\, n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}. \tag{2.3}$$

The moments of the binomial distribution follow easily from differentiating the generating function

$$G(z, N) = (pz + q)^N = \sum_{k=0}^{N} p(k, N)\, z^k \tag{2.4}$$

with respect to z, then setting $z = 1$. If we denote the stochastic variable representing "success in N trials" by S_N, then,

$$< S_N >= N\, p, \quad \text{and the dispersion is } \sigma^2 \equiv < S_N^2 > - < S_N >^2 = N\, p\, q\,. \tag{2.5}$$

Both these results, that the mean value is "obvious," and that the width of the distribution is proportional to the square-root of N, are famous, and important.

Poisson Statistics

Binomial statistics has two important limits, obtained when the number of trials is large, $N >> 1$. The first, which leads to Poisson statistics, is the limit in which $N \to \infty,\ p \to 0$ in such a manner that the mean value, $N\,p$, remains constant. We denote that limit by λ. If we write $P(k,N)$ as

$$P(k, N) = \frac{1}{k!} N(N-1)...(N-(k-1))\, p^k\, [\,(1-p)^{\frac{1}{p}}\,]^{(N-k)p}$$

and pass to the limit, we easily get

$$P(k, N) \to \frac{1}{k!} \lambda^k\, e^{-\lambda} = P(k, \lambda) \tag{2.6}$$

which is the Poisson distribution for k successes. Its normalization is obvious, and both its mean and dispersion are equal to λ, the single parameter that characterizes the distribution. The generating function is

$$G(z, \lambda) = e^{(z-1)\lambda} = \sum_{k=0}^{\infty} P(k, \lambda)\, z^k\,. \tag{2.7}$$

There are many nice examples of the Poisson distribution. We do not always, formally, carry out the limiting process. We use the distribution when we meet situations where N is very large and p very small. Perhaps the simplest example is the image of many independent archers letting fly at a difficult target. Suppose that experience shows that, on average, there are λ hits per barrage. Then, the probability of k hits is $P(\,k, \lambda\,)$.

Other examples:

1) Beer's Law: A beam of particles (e.g., photons or neutrons) passes through a slab of material of thickness *L*. Let the material be divided into a large number of slices, $N = L/\Delta x$, arranged perpendicular to the beam. Suppose there is a small probability, proportional to the slab thickness, $p = \mu \Delta x$, that an encounter in the slab will remove a particle from the beam. This is the Poisson situation, with $\lambda = N p = \mu L$. Then, for example, the probability of traversing the slab without deflection is $\exp(-\mu L)$ (Beer). If, further, the encounter can be viewed as caused by a collision with one of many tiny, identical scatterers (or absorbers), whose number density is *n*, the parameter μ would be equal to $n\sigma$, σ being the *cross section* for the process. An obvious generalization of Beer's Law to a material of gently varying properties, $\mu(x)$, alters the exponential, so that

$$\mu L \rightarrow \int_0^L dx\, \mu(x).$$

The new expression is the *optical path length.*

2) Reception of Signals: Let a receiver be operated for temporal periods *T*, during which, on average, λ distinct signals are heard. Let the interval be divided into many small intervals Δt, during which the probability of receiving a signal is independent of all else. The probability is small, proportional to Δt. Then, as above, Poisson statistics prevail in the interval (0, *T*), the probability of no signals, for example, is $\exp(-\lambda)$. This, in turn, is an example of a *Poisson Time Series*, or *Process*, in which a sequence of pulses might signify the random turning on of a switch, or perhaps a random sequence of births—all independent events, of course.

3) Radioactive Decay: A simple theory of, say, alpha-particle radioactivity would picture the particle rattling back and forth in its nuclear potential, challenging the barrier ω times per second, in the hope of escape. In time *t*, there are many encounters, ωt. With each attempt to escape (penetrate) we associate a tiny tunneling probability, α. Then, Poisson prevails with $\lambda = \alpha\omega t$, and we have the famous law of radioactive decay, survival = $\exp[-(\alpha\omega)t]$.

4) Number Fluctuations: An ideal gas, composed of *N* point-particles, occupies a volume V. What is the probability that *k* particles will be found in a tiny volume ΔV? Again, we have *N* independent attempts with success $p = \Delta V / V$. The mean number of particles resident will be $\lambda = (N/V)\,\Delta V$. And the chance of our encountering vacuum is the astronomically small $\exp[-(N/V)\,\Delta V]$. We might estimate the fluctuation in number (or density) by the ratio of the square root of the dispersion to the mean number. That ratio is $1/\sqrt{\lambda}$.

Gaussian Statistics

Again we consider the case $N >> 1$, so that the binomial distribution, which is discrete, can be replaced by one that is smooth. Here, we place no conditions on the probability of individual events but, rather, remain near the center of the distribution, in a sense that will be specified. We will miss fine details of the "wings" of the distribution. The detailed calculation will be done for $p = q = ½$ .The general result will be given, without proof.

Central to our calculation is the Stirling approximation to the factorial

$$N! \sim \sqrt{2\pi N}\left(\frac{N}{e}\right)^N \qquad N >> 1. \tag{2.8}$$

We wish to evaluate the probability for k successes $P(k,N) = \frac{1}{2^N}\binom{N}{k}$ when N is very large, and k is near the maximum, which occurs at $k=N/2$. Thus we want to estimate

$$\frac{1}{2^N}\binom{N}{\frac{N}{2}+\ell} = \frac{1}{2^N}\frac{N!}{(\frac{N}{2}+\ell)!\,(\frac{N}{2}-\ell)!} \qquad N,\ell >> 1. \tag{2.9}$$

When we use the Stirling approximation here, the factors e^N cancel and we find

$$\frac{1}{2^N}\binom{N}{\frac{N}{2}+\ell} = \frac{1}{2^N}\frac{1}{\sqrt{2\pi}}\frac{N^{N+\frac{1}{2}}}{(\frac{N}{2}+\ell)^{\frac{N}{2}+\ell+\frac{1}{2}}\ (\frac{N}{2}-\ell)^{\frac{N}{2}-\ell+\frac{1}{2}}} \qquad N,\ell >> 1$$

or

$$= \sqrt{\frac{2}{\pi N}}\frac{(1-\frac{2\ell}{N})^{\ell}}{(1+\frac{2\ell}{N})^{\ell}\ (1-(\frac{2\ell}{N})^2)^{\frac{N}{2}}} \qquad N,\ell >> 1.$$

Now we describe the limit more carefully. N,ℓ will become large in such a manner that $\frac{\ell^2}{N}$ remains finite, fixed. Then we find

$$(1\pm\frac{2\ell}{N})^{\ell} \to \exp[\pm\frac{2\ell^2}{N}] \text{ and } (1-(\frac{2\ell}{N})^2)^{\frac{N}{2}} \to \exp[-\frac{2\ell^2}{N}] \text{ or}$$

$$\frac{1}{2^N}\binom{N}{\frac{N}{2}+\ell} \sim \sqrt{\frac{2}{\pi N}}\ \exp[-\frac{2\ell^2}{N}] \qquad N, \ell >> 1,\ \frac{\ell^2}{N}\ \text{fixed}$$

and,

$$P(k, N) \sim \sqrt{\frac{2}{\pi N}}\ \exp[-\frac{2}{N}(k-\frac{N}{2})^2]\ , \tag{2.10}$$

which is properly normalized if *k* ranges over (+/-) infinity. It is no surprise to find that the general case, associated with a bit more computation, yields

$$P(k, N) \sim \frac{1}{\sqrt{2\pi N\ pq}}\ \exp[-\frac{1}{2N\ pq}(k-Np)^2]\ . \tag{2.11}$$

As to normalization, to the extent that one can replace sums by integrals, we have

$$\sum_{k=0}^{N} P(k,N) \sim \int_0^N dk\ \sqrt{\frac{2}{\pi N}}\ \exp[-\frac{2}{N}(k-\frac{N}{2})^2] = \sqrt{\frac{2N}{\pi}}\ \int_0^1 dx\ \exp[-2N(x-\frac{1}{2})^2]\ .$$

When *N* is large, the limits of the last integral can be replaced by $\pm\infty$ with negligible error, whereupon the last term is unity.

The Gaussian has so many interesting properties that we shall devote an entire lecture to it.

Lecture 3

Gaussians

The Gaussian, or normal distribution, appears to pervade stochastics. It appears in medical, financial, environmental studies. Its ubiquity rests on its remarkable properties. Most important of these is its appearance as a limiting distribution.

The first of many interesting properties of the Gaussian is that compounding two (or more) Gaussians gives, again, a Gaussian. More precisely, we ask: "If stochastic variables X_1, X_2 are Gaussian, with values ranging $\pm\infty$, and are statistically independent, what is the distribution of $Y = X_1 + X_2$? " To get *p(y)* we use a device noted earlier, namely

$$p(y) = \int dx_1 \int dx_2 \; p(x_1, x_2) \; \delta(x_1 + x_2 - y) \,, \tag{3.1}$$

with $p(x_1, x_2)\, dx_1\, dx_2$ the probability of obtaining x_1, x_2 in the interval $dx_1\, dx_2$. By hypothesis, $p(x_1, x_2) = p_1(x_1)\; p_2(x_2)$, and each of the factors is a Gaussian, with its own mean, a_1, a_2, and dispersion. The integration, which is normalized in the variable *y*, selects those values of x_1, x_2 that contribute to the sum. Upon integration we have

$$p(y) = \frac{1}{2\pi\sigma_1\sigma_2} \int dx_1 \exp\{-\frac{1}{2\sigma_1^2}(x_1 - a_1)^2 - \frac{1}{2\sigma_2^2}(y - (a_1 + a_2) - (x_1 - a_1))^2\} \tag{3.2}$$

whereupon straightforward algebra ("completion of the squares, etc.") brings us to

$$p(y) = \frac{1}{\sqrt{2\pi}\,\sigma_3} \exp[-\frac{(y - (a_1 + a_2))^2}{2\sigma_3^2}] \qquad -\infty < y < \infty \tag{3.3}$$

with $\sigma_3^2 = \sigma_1^2 + \sigma_2^2$. Thus, expectations add, and variances add. If we add N identical Gaussians, the width of the resulting Gaussian is to its mean as $1/\sqrt{N}$. That sums of Gaussians and, in special cases, integrals over continuously indexed Gaussians are also Gaussian turns out to be of great importance in applications.

One might ask: Are there other distributions with the property that the distribution of $Y = X_1 + X_2$, resembles, or repeats, that of the *X's*—given that X_1, X_2 are independent and (statistically) identical? These are called *stable* distributions. To find them, we look to the characteristic functions. If we replace the delta function in Eq. (3.1) by its representation,

$$\delta(x_1 + x_2 - y) = \frac{1}{2\pi}\int dk\ \exp[-ik(x_1 + x_2 - y)]\ , \tag{3.4}$$

and do some easy integrations, we find $C_Y(k) = C_X(k)^2$. For stability, we should search for distributions whose characteristic function is a simple exponential. The simplest of these is $C(k) = \exp[-c|k|^\alpha]$, $0 < \alpha \le 2$, which generates a normalized distribution function

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx}\ \exp[-c|k|^\alpha], \tag{3.5}$$

symmetrical about $x = 0$. It produces the Gaussian $(\sigma^2 = 2c)$ when $\alpha = 2$, and the Cauchy distribution $p(x;c) = \frac{1}{\pi}\frac{c}{c^2 + x^2}$, when $\alpha = 1$. One can show that the wings of the distribution behave as $p(x;\alpha) \sim \frac{const}{|x|^{1+\alpha}}$, in the non-Gaussian case. Thus, these symmetric distributions have no moments. The *scaling* property of the distribution follows from consideration of $p(Ax;\ c,\ \alpha)$. Thus,

$$p(Ax;\ c,\ \alpha) = \frac{1}{A}\,p\left(\frac{x}{A};\ \frac{c}{A^\alpha}, \alpha\right). \tag{3.6}$$

The class of stable distributions may be extended from the centered, symmetric functions discussed above, to shifted, skewed distributions. Shifting by an amount μ is achieved by adding $-ik\mu$ to the exponent, and skewness is introduced by a new parameter, β. Then, the general expression for the class of *Levy-Stable* distributions is

$$A(k) = Exp[-ik\mu - c\,|k|^\alpha\ (1 + i\beta\,\mathrm{sgn}(k)\,\Phi(\alpha,k)\,]] \tag{3.7}$$

with $\Phi(\alpha,k) = \tan\left(\frac{\alpha\pi}{2}\right)$, $\alpha \neq 1$, $\Phi(\alpha,k) = -\frac{2}{\pi}\log|k|$, $\alpha = 1$.

But there is more to the Gaussian than its stability. It can be generated by non-Gaussian distributions. This brings us to

Random Flights and the Central Limit Theorem

Let us generalize our addition $Y = X_1 + X_2$ to the compounding of *N* independent "flights," each governed by the same distribution, but not necessarily Gaussian. Then—and we work in one dimension−the distribution of the resultant is

$$P(y,N) = \int dx_1 \int dx_2 \ldots \int dx_N\ p(x_1)\,p(x_2)\ldots p(x_N)\ \delta\left(y - \sum_{m=1}^{N} x_m\right). \tag{3.8}$$

We progress by introducing the Fourier representation of the delta-function, which gives, quickly,

$$P(y,N)=\frac{1}{2\pi}\int dk\, e^{i\,k\,y}[\,p_{\#}(k)]^N$$

with (3.9)

$$p_{\#}(k)=\int dx\, e^{-i\,k\,x}\,p(x)\,.$$

Thus, the characteristic function for the resultant is the *N* th power of the characteristic function for the individual flight. Suppose that the flights are symmetrical, whereupon *p(x)* is even and odd powers of *k* will be absent. Consider

$$[\,p_{\#}(k)]^N=[1-\frac{1}{2!}\mu_2\,k^2+\frac{1}{4!}\mu_4 k^4-...]^N$$

$$=[1-\frac{1}{N}\{\frac{1}{2!}\mu_2\,Nk^2+\frac{1}{4!}\frac{\mu_4}{N}(Nk^2)^2-...\}]^N \qquad (3.10)$$

in the limit of *N* large, k^2 small and their product fixed. (Thus we are considering the large-displacement behavior after many steps.) In that limit,

$$[\,p_{\#}(k)]^N\rightarrow\exp[-\frac{1}{2}\mu_2\,Nk^2\,] \qquad (3.11)$$

and *P(y, N)* has the form, under our assumptions, of a Gaussian with variance $N\mu_2$. This result suggests the famed Central Limit Theorem, which describes the extent to which the distribution of the sum of *N* independent events may be considered a Gaussian, overall.

Having noted that the sum of Gaussians is distributed in a Gaussian manner, it is natural to inquire about the integral of Gaussians, the "sum" of an infinity of such functions. For example, $Y=\int_0^1 dt\;X(t)\Leftrightarrow Lim\sum_{n=1}^{N}\Delta t\;X_n$ with $X_n\equiv X(t_n)$. We'll begin with the simple case in which the X_n may be regarded as uncorrelated, identically distributed, centered Gaussians. (We relax the condition on correlation later in this lecture.) If we assume that *Y* is, indeed, Gaussian, the dispersion of *p(y)* will be

$$<Y^2>=Lim\sum_{n=1}^{N}\Delta t\sum_{n'=1}^{N}\Delta t<X_n\,X_{n'}>\rightarrow\int_0^1 dt\int_0^1 dt_1<X(t)\,X(t_1)>\,. \qquad (3.12)$$

But this would appear to vanish, since $<X_n\,X_{n'}>=<X_n><X_{n'}>=0$. A troubling result, but salvation is possible—in a special case. Since $<X_n^2>\neq 0$, we might consider $<X(t)\,X(t_1)>=\Gamma\,\delta(t-t_1)$, $<Y^2>=\Gamma$ as the limiting correlation. More insight may be

obtained from the *Wiener-Khinchin* relation, which is deduced in Appendix B. (The discussion in the appendix relies somewhat on material treated in Lecture 5.) The relation asks that we consider the frequency analysis of the correlation function, and also regard the local intensity (dispersion) as compounded out of contributions from many frequencies. If the situation is stationary and homogeneous (no boundary effects), we may write:

$$< X(\tau)X(0) >= \frac{1}{2\pi}\int d\omega \exp(i\omega\tau) I(\omega) \ , \ < X^2 >= \frac{1}{2\pi}\int d\omega I(\omega). \quad (3.13)$$

The δ-function correlation refers, then, to $I(\omega) = 1$, a uniform distribution of intensity on the frequency scale. This situation, this $\delta - correlated\ process$, is referred to as *white noise.* In our case it is *Gaussian white noise.* Of course, since $I(\omega) = 1$ implies a divergent integral, one must use this model carefully, as a limiting case, applying "cut-offs" when needed, with care. Going further, one might argue that our discussion may be extended to integrals of local variables *X(t)* whose distributions are not Gaussian, but qualify via the Central Limit Theorem. Finally, we will, ahead, relax the condition of white noise. We shall find that the integral of correlated Gaussians is indeed Gaussian.

As a first example of Gaussian noise, we'll consider a problem in the physics of clouds, or

The Propagation of Radiation through a Random Medium

Recall that we applied Poisson statistics to the interaction of radiation with a homogeneous slab of material, taking the probability of interaction with a slab of thickness Δx to be $\mu\Delta x$. We deduced that the probability of transmission, of emerging unscathed from a slab of thickness *L*, was $T(L) = Exp[-\mu L]$. Suppose now that the slab is not homogeneous, a cloud, perhaps, and that we have a complicated, only partially known $\mu(x)$, leading to $T(L) = Exp[-\int_0^L dx\,\mu(x)]$. (The integral is known as the "optical depth.") We measure the transmission many times, obtaining slightly different values—fluctuations about a mean—and ask whether statistical methods might be applied.

The $\mu(x)$ will be thought of as a continuous infinity of independent, identical random variables each fluctuating about an *x*-dependent average $< \mu >(x)$. In this simplest case, the average is independent of position, and is denoted by $< \mu > \equiv \mu_0$.

Thus, $\mu(x) = \mu_0 + M(x)$, *M(x)* being stochastic and well-enough behaved that the Central Limit Theorem holds. Then, the integral, *I(L)*, will be Gaussian, with mean value zero. Its distribution, described by

$$p(I;\sigma_I^2) = \frac{1}{\sqrt{2\pi}\,\sigma_I} \exp[-\frac{I^2}{2\sigma_I^2}] \ ,$$

is determined by a single quantity, the variance.

The fluctuating transmission, $T = e^{-\mu_0 L}\, e^{-I} \;\; = e^{-\mu_0 L}\; T_1$, is described by the probability distribution

$$P(T) = e^{\mu_0 L}\; p_1(T_1)$$

$$p_1(T_1) = \int dI\; p(I;\sigma_I^2)\; \delta(T_1 - e^{-I}) \qquad 0 < T_1 < \infty \tag{3.14}$$

or
$$p_1(T_1)\,|\,\frac{dT_1}{dI}\,| = p(I_*), \quad I_* = \log(1/T_1)\ . \tag{3.15}$$

(Of course, since the Gaussian gives non-zero probability for un-physical events—e.g., negative densities—our model must be treated with caution.) The integral, Eq. (3.14) gives

$$p_1(T_1) = \frac{1}{T_1}\, p\,(\log(1/T_1))$$

and, finally, we obtain the *log-normal* distribution,

$$P(T) = \frac{1}{\sqrt{2\pi}\,\sigma_I\, T}\, \exp[-\frac{(\log T_1)^2}{2\sigma_I^2}] \tag{3.16}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_I\, T}\, \exp[-\frac{(\log(1/T) - \mu_0 L)^2}{2\sigma_I^2}]\ .$$

The last form shows clearly that as the stochasticity diminishes, the variance vanishing, the distribution will approach delta-function form and yield $T \to e^{-\mu_0 L}$.

The *log-normal* distribution would appear to give palpable probability to large fluctuations. Its moments are

$$\mu_n \equiv \int dT\, T^n\, P(T\,) = e^{-n\mu_0 L} \int dI\; p(I;\sigma^2)\; e^{-nI} \;= \exp[-n\mu_0 L + \frac{1}{2} n^2 \sigma_I^2]. \tag{3.17}$$

Should we have white noise, $< M(x) M(x_1) > = \Gamma\, \delta(x - x_1)$, $\sigma_I^2 = \Gamma\, L$ and

we find
$$\mu_n = \exp[-n\mu_0 + \frac{1}{2} n^2\, \Gamma\,]L\ .$$

The moments appear to be large. But further consideration suggests that in realistic cases the dispersion is, relatively, very small. Then we may take seriously the conclusion that fluctuations tend to increase the likelihood of transmission.

Another example of the use of Gaussian white noise is one of considerable historical importance. It is *Brownian motion*, the motion of a particle immersed in a fluid and bombarded by tiny but effective collisions with the molecules constituting the fluid. We will devote Lecture 4 to its exposition.

The *N*-dimensional Gaussian

We have discussed objects which are the sum of a large number of uncorrelated Gaussians. We can broaden the subject by examining Gaussians in *N* variables which are correlated. Thus, consider the probability distribution

$$G_N(x_1,...x_N) = C_N \exp[-\frac{1}{2}\sum_{m,n=1}^{N} B_{mn} x_m x_n], \tag{3.19}$$

where the sum is a real, positive definite, quadratic form. The normalization is achieved by introducing a change of variable—an orthogonal transformation—(a rotation) that diagonalizes B_{mn}. Thus, $x_m = \sum S_{mp} y_p$ and the matrix S and its transpose cause the product $S^T B S$ to be diagonal, displaying eigenvalues β_m. Since the Jacobian of the transformation is unity, we have

$$\int dx_1 ... \int dx_N \exp[-\frac{1}{2}\sum_{m,n=1}^{N} B_{mn} x_m x_n] = \int dy_1 ... \int dy_N \exp[-\frac{1}{2}\sum_{m,n=1}^{N} \beta_m y_m^2] \ . \tag{3.20}$$

The integrals are easy, whereupon the normalized distribution is:

$$G_N(x_1,...x_N) = \frac{\sqrt{Det\, B}}{(2\pi)^{\frac{N}{2}}} \exp[-\frac{1}{2}\sum_{m,n=1}^{N} B_{mn} x_m x_n] \ . \tag{3.21}$$

The X_m are correlated, and the matrix of their correlations, $< X_m \ X_n >$, is particularly interesting. We get it by first considering the diagonal matrix

$$< Y_r Y_s > = \frac{\sqrt{Det\, B}}{(2\pi)^{\frac{N}{2}}} \int dy_1 ... \int dy_N \ y_r \ y_s \ \exp[-\frac{1}{2}\sum_{m,n=1}^{N} \beta_m y_m^2], \tag{3.22}$$

which is simply $< Y_r Y_s >= \frac{1}{\beta_s}\delta_{rs}$. But the rhs is equal to $S^T B^{-1} S$, while the lhs is seen to be $< Y_r Y_s >= \sum_{k\,m} (S^T)_{rk} < X_k X_m > S_{ms}$. (We use $S^T = S^{-1}$.)

Thus, $< X_m X_n > \equiv C_{mn} = B^{-1}_{mn}$. The matrix of correlations, the *covariance matrix*, is the inverse of the matrix that defines the Gaussian—a surprising and pleasant result. In fact, one can show by means similar to those used above, that the characteristic function for the *N*-Gaussian has the simple form

$$C_N(k_1,...k_N) = Exp[-\frac{1}{2}\sum_{m,n=1}^{N} C_{mn}\ k_m\, k_n\ ]. \tag{3.23}$$

Should the correlation matrix be diagonal, the *N*-Gaussian becomes the product of independent Gaussians, and vice-versa.

A final result concerns higher moments of the N-Gaussian,

$$< X_i X_j X_l X_m .... > = \int dx_1 ... \int dx_N\ \ x_i\, x_j x_l x_m ...\ \ \ G_N(x_1,...x_N)\,.$$

Recall that

$$\int dx_1 ... \int dx_N\ \ \ Exp[-i\sum_{n=1}^{N} k_n\, x_n\,] G_N(x_1,...x_N)$$

$$= Exp[-\frac{1}{2}\sum_{m,n=1}^{N} C_{mn}\ k_m\, k_n\ ]. \tag{3.24}$$

The moments will be expressed in terms of C_{mn} after we expand both exponentials in Eq. (3.24) and compare terms. We see easily that if a moment contains an odd number of factors, it will vanish. If it contains 2*n* factors it will be expressed by *n* factors of form $C_{km} = < X_k X_m >$. It is no surprise, then, that

$$< X_i X_j X_l X_m .... >= \sum < X_s X_t >< X_p X_v > ... \tag{3.25}$$

where the sum on the rhs requires that we sum over all pairings of the indices appearing on the lhs. For example,

$$< X_1 X_2 X_3 X_4 >= < X_1 X_2 >< X_3 X_4 > + < X_1 X_3 >< X_2 X_4 > + < X_1 X_4 >< X_2 X_3 >\,.$$

A little counting convinces one that there are $1 \cdot 3 \cdot 5 \cdot 7 ... (2n-1)$ terms in the expression. This result is of considerable importance when one considers, for example, the multiple scattering of waves in disordered media.

The Integral of a Gaussian

We return to this interesting issue, now taking the $X_n \equiv X(t_n)$ of $Y = \int_0^1 dt\ X(t) \Leftrightarrow Lim \sum_{n=1}^{N} \Delta t\ X_n$ to be the stochastic variables represented by the $p(x_1,...x_N)$ of Eq. (3.21). Then

$$p(y) = \int dx_1 ... \int dx_N\ p(x_1,...x_N)\ \delta(y - \sum_{n=1}^{N} \Delta t\ X_n)\,. \tag{3.26}$$

We use the integral representation of the delta-function, and recognize the characteristic function, to arrive at

$$p(y) = \int \frac{dy}{2\pi} e^{i k y}\ C_N(k_1,...k_N) \qquad \text{with each } k_m \to k\,\Delta t\,.$$

$C(k_1,...k_N)$ has become $Exp[-\frac{1}{2}k^2\sigma^2]$ with $\sigma^2 = \sum_{m,n=1}^{N} (\Delta t)^2 C_{mn}$. It is Gaussian and in all but pathologic cases will permit smooth passage to the limit $N \to \infty,\ \Delta t \to 0,$ $N\Delta t \to 1$, to give

$$p(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp[-\frac{y^2}{2\sigma^2}] \tag{3.27}$$

with

$$\sigma^2 = \int_0^1 dt \int_0^1 dt_1\ C(t,t_1)\,. = \int_0^1 dt \int_0^1 dt_1 < X(t) X(t_1) >\,.$$

Gaussian white noise, considered earlier, is generated by $C(t,t_1) = \Gamma\,\delta(t - t_1)$. It is common to incorporate a scale, T, over which correlations are important and to represent the correlation function as

$$C(t,t_1) = \frac{\Gamma}{T} Exp[-\frac{|\,t - t_1\,|}{T}]\ . \tag{3.28}$$

It should be obvious that our argument may be extended to go beyond simple integrals, to demonstrate the Gaussian nature of more general linear functionals, of form

$$Y(t) = \int_0^t dt_1\ K(t,t_1)\,X(t_1) \tag{3.29}$$

for example. We will be using this result when we consider Brownian motion.

Some Thermodynamics

The Gaussian of many variables appears often in thermodynamics and statistical mechanics. A classic example is its use in the analysis of fluctuations—fluctuations at the meso-and macroscopic level, those barely perceptible, and those occurring at thermodynamic *critical points*. One of the earliest treatments, directed at critical opalescence, was given by Albert Einstein in 1910.

All revolves about Boltzmann's famous expression for the entropy of a system at equilibrium, $S_0 = k_B \log W_0$ where W_0 is a probability assigned to the macroscopic, thermodynamic state. Einstein inverted the expression to read $W = Exp(S/k_B)$ and applied it to states out of equilibrium, regarding *W* as the probability that an off-equilibrium state, characterized by a smaller entropy, would occur as a fluctuation. Suppose the system is described by a set of variables, $X_1, X_2, \ldots$, the entropy being $S(X_1, X_2, \ldots)$. Assume that states out of equilibrium are described by these variables too. The entropy of these will be written $S = S_0 - \frac{1}{2}\sum_{m,n=1}^{N} B_{mn}(X_m - X_m^0)(X_n - X_n^0) + \ldots$ with $B_{mn} = -(\frac{\partial^2 S}{\partial X_m \partial X_n})_0$, the expression describing modest fluctuations. (The variables describing the fluctuations are set equal to zero in the equilibrium state. They might, for example, be the Fourier amplitudes of disturbances in density.) The first derivatives in the Taylor expansion are absent, because the entropy is presumed to be maximum at equilibrium. Thus, we have a distribution of the form

$$W(x_1, x_2, \ldots) = C\, Exp[-\frac{1}{2k_B}\sum_{m,n=1} B_{mn} x_m x_n] \qquad (3.30)$$

for the probability of a particular fluctuation. The various correlations, being elements of the matrix inverse of B_{mn} , may be expressed via thermodynamic derivatives, which refer, of course, to the equilibrium state.

A word of caution: Gaussians, attractive in many ways, can be over-used. A common attitude is, "When uncertain, take the distribution to be Gaussian; after all, there's the Central Limit Theorem and all that." But ultimately, when a complicated physical, biological, or financial system is being treated stochastically, the issue is truth rather than convenience. The Gaussian emphasizes small deviations from the mean; its wings are relatively "thin" and surprises are unlikely. In reality, this may not be the case at all.

Lecture 4

Gaussian Statistics and Brownian Motion

We are often called upon to analyze a "noisy" system. The system may come from the world of physics, of engineering, of biology, or finance. Its inherent dynamics may be linear or (God forbid) non-linear. There may or may not be a nice separation between the time scales intrinsic to the system and the time scale of the noise.
Clearly, we face a rich and robust challenge.

The simplest non-trivial model available for study is that of *Brownian Motion;* the tool is the *Langevin Equation*.

We consider a particle, small but macroscopic, immersed in a fluid. It undergoes rapid collisions with the molecules which constitute the fluid and, perhaps, it is influenced by an external field, too. The collisions constitute both a rapidly fluctuating force, *F(t),* whose average is zero, and a viscous "drag" that damps the motion of the particle. Thus, we might describe the motion through a *Langevin Equation*,

$$\frac{dv(t)}{dt} + \beta\, v(t) = \frac{1}{M}[F(t) + F_{ext}] = L(t) + L_{ext}\,, \tag{4.1}$$

with integral

$$v(t) = v_0\, e^{-\beta t} + \int_0^t dt' e^{-\beta(t-t')}[L(t') + L_{ext}] \tag{4.2}$$

describing the motion of a Brownian particle introduced with velocity v_0 into the background fluid. Typically, the damping co-efficient is $\beta = \frac{6\pi\eta a}{M}$ where η is the viscosity of the host fluid and *a* is the characteristic size of the particle—here taken to be a sphere. (The coefficient may depend upon the particle-fluid boundary interaction, as well.) What is crucial is that the time associated with the damping, $1/\beta$, is great in comparison with the characteristic time for fluctuation of the collisional force—the collision time. Thus, the time-average of *L(t)* over an interval $1/\beta$ is essentially zero.

In our analysis we treat the *L(t)* as stochastic variables, as centered Gaussians, $<L(t)> = 0$, at each time. Further, $< L(t)\, L(t') > = 2\gamma\,\delta(t - t')$ and we will be dealing with Gaussian white noise. Much of the argument resembles that of the example discussed in Lecture 3, with *x* replaced by *t'* and *L* by *t.* The velocity *v(t)* has become a stochastic variable. We replace this time-dependent variable by a continuous infinity of random variables *V(t)* which will inherit the stochasticity of *L(t)*.

We proceed with the stochastic version of Eq. (4.2) neglecting the external force. Then, averaging over the values of *L* gives

$$< V(t) >= v_0 e^{-\beta t} . \tag{4.3}$$

It is useful to work with the fluctuation,

$$V_1(t) = V(t) - v_0 e^{-\beta t} = \int_0^t dt' e^{-\beta(t-t')} L(t') . \tag{4.4}$$

Since the integral is centered Gaussian at each instant of time, we conclude that $V(t)$ is Gaussian too, with average $v_0 e^{-\beta t}$. The variance of $V(t)$ is $< V_1(t)^2 >$, for the distribution of the two *V's* are related merely by a time-dependent shift. Then, given that $v = v_0$ at $t=0$, the probability of observing *v* at time *t* (compare *I(L)* and *p(I)* in the problem of Lecture 3) is

$$P(v,t|v_0,0) = \frac{1}{\sqrt{2\pi}\,\sigma(t)} \exp[-\frac{(v - v_0 e^{-\beta t})^2}{2\sigma(t)^2}] \tag{4.5}$$

$$\text{with } \sigma(t)^2 = < V_1(t)^2 > .$$

Computation gives

$$< V_1(t)^2 >= \int_0^t dt' \int_0^t dt'' \, e^{-\beta(t-t')} \quad e^{-\beta(t-t'')} < L(t')L(t'') > = \frac{\gamma}{\beta}(1 - e^{-2\beta t}) . \tag{4.6}$$

Further integration gives the statistics of position, thus

$$X_1(t) = \int_0^t dt_1 V_1(t_1) = X(t) - (x_0 + \frac{v_0}{\beta}(1 - e^{-\beta t})) \tag{4.7}$$

$$= \int_0^t dt_1 \int_0^{t_1} dt_2 \, e^{-\beta(t_1 - t_2)} L(t_2)$$

$$= \frac{1}{\beta} \int_0^t dt_1 (1 - e^{-\beta(t-t_1)}) L(t_1) .$$

$X_1(t)$ is the integral of a δ-correlated Gaussian. It is Gaussian, with zero average. The average of *X(t)* is $x_0 + \frac{v_0}{\beta}(1 - e^{-\beta t})$, the deterministic trajectory. The variance of $X_1(t)$ is

$$< X_1(t)^2 >= \sigma_x(t)^2 = (\frac{2\gamma}{\beta})^2 \int_0^t dt_1 \, (1 - e^{-\beta t_1})^2 \ , \tag{4.8}$$

$$= \frac{\gamma}{\beta^3}(2\beta t - 3 + 4e^{-\beta t} - e^{-2\beta t}) .$$

The variance is of course independent of (x_0, p_0). The probability distribution for displacement is, then,

$$P(x,t\,|\,x_0,v_0,0) = \frac{1}{\sqrt{2\pi}\sigma_x(t)}\ \exp[-\frac{(\,x-(x_0+\frac{v_0}{\beta}(1-e^{-\beta t}\,))\,)^2}{2\sigma_x(t)^2}]\,. \qquad (4.9)$$

The distribution is peaked about the deterministic trajectory, its width growing with time. For sufficiently long times, information about the initial state is forgotten. When βt is large, $\sigma(t)^2 \rightarrow \frac{2\gamma}{\beta^2}t = 2Dt$, characteristic of simple diffusion. Indeed, we note

$$P(x,t\,|\,x_0,v_0,0)\ \rightarrow\ P(x,t) = \frac{1}{\sqrt{4\pi D t}}\ \exp[-\frac{x^2}{4Dt}]\,, \qquad (4.10)$$

which is a famous expression. The connection between *D*, the measure of a dissipative process, and γ, the measure of a fluctuating force, is an example of a *fluctuation-dissipation* theorem.

The full conditional probability, $P(x,v,t|x_0,v_0,0)$, is Gaussian in two variables, $x_1 = x-(\,x_0+\frac{v_0}{\beta}(1-e^{-\beta t})\,)$ and $v_1 = v - v_0 e^{-\beta t}$. These describe deviation from the deterministic trajectory. Thus,

$$P(x,v,t|x_0,v_0,0) = \frac{\sqrt{Det\,B}}{(2\pi)}\ \exp[-\frac{1}{2}(B_{11}\,x_1^2 + B_{12}x_1v_1 + B_{22}v_1^2)] \qquad (4.11)$$

the *B*-matrix being the inverse of the matrix of correlations,

$$C_{11} = < X_1(t)X_1(t) >,\ \ C_{12} = C_{21} = < X_1(t)V_1(t) >,\ \ C_{22} = < V_1(t)V_1(t) >\,.$$

The marginal distributions which we have noted earlier are then

$$P(v,t|v_0,0) = \int dx\ \ P(x,v,t|x_0,v_0,0)\ \ \text{and}\ \ P(x,t\,|\,x_0,v_0,0) = \int d\,v\ P(x,v,t|x_0,v_0,0)\,. \qquad (4.12)$$

We've encountered C_{11} and C_{22}, while

$$C_{12} = \frac{1}{\beta}\int_0^t dt_1\,(1-e^{-\beta(t-t_1)}\,)\int_0^t dt_2\,e^{-\beta(t-t_2)} < L(t_1)L(t_2) >$$

$$= \frac{\gamma}{\beta^2}(1-e^{-\beta t})^2\,. \qquad (4.13)$$

The last result may be surprising at first sight, for one might expect $X_1(t)$ and $V_1(t)$ to become un-correlated at long times, with the consequent vanishing of C_{12}. In fact, if we note that

$$< X_1(t)V_1(t) > = < X(t)V(t) > - < X(t) >< V(t) > = \frac{1}{2}\frac{d}{dt} < X^2(t) > - < X(t) >< V(t) >$$

and that $< X^2(t) > \to 2Dt$, $< V(t) > \to 0$ at long time intervals, we find, indeed that $C_{12} \to D = \frac{\gamma}{\beta^2}$. The fluctuations remain correlated.

Generalizations abound in the world of linear systems. For example, the noise-driven harmonic oscillator, or the RLC electrical circuit, etc., may be treated in a similar manner. The solution will be of the form

$$X(t) = x_0\ f_1(t) + v_0\ f_2(t) + \int_0^\infty dt'\ G(t,t')\,L(t') \tag{4.14}$$

$$= \ x(x_0, v_0, t) + \int_0^\infty dt'\ G(t,t')\,L(t')$$

with $f_1(0) = 1$, $f_1'(0) = 0$, $f_2(0) = 0$, $f_2'(0) = 1$. Note that $x(x_0, v_0, t)$ is the deterministic trajectory. The Green's function is

$$G(t,t') = 0 \quad t < t' \qquad G(t,t') = g(t-t') \quad t > t' \tag{4.15}$$

and we conclude that

$$P(x,t \mid x(0), v(0), 0) = \frac{1}{\sqrt{2\pi}\sigma(t)}\ \exp[-\frac{(x - x(x_0, v_0, t))^2}{2\sigma(t)^2}] \tag{4.16}$$

$$\sigma(t)^2 = 2\gamma \int_0^t dt'\, g(t-t')^2 .$$

As an example, the noise-driven, damped harmonic oscillator, with $\beta_1^2 = \beta^2 - 4\omega^2 > 0$ yields

$$\begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix} = e^{-\frac{1}{2}\beta t} \begin{pmatrix} \cosh\frac{\beta_1 t}{2} + \frac{\beta}{\beta_1}\sinh\frac{\beta_1 t}{2} \\ \frac{2}{\beta_1}\sinh\frac{\beta_1 t}{2} \end{pmatrix} \tag{4.17}$$

and

$$g(t-t') = \frac{2}{\beta_1} e^{-\frac{1}{2}\beta(t-t')} \sinh\frac{\beta_1(t-t')}{2} \qquad t>t' . \tag{4.18}$$

We must point out somewhere in this lecture a logical difficulty that occurs in dealing with the Langevin Equation. We may encounter expressions like $\int dt\, L(t)\, v(t)$, which are ill-defined. *L(t)* may consist of a sequence of un-correlated spikes $\delta(t-t_i)$. These, in driving the equation for $v(t)$, induce jumps in *v(t)* at t_i and hence ambiguity in the evaluation of the integral. The issue may be resolved by treating the delta function as the limit of a proper function (Stratonovich) or by inventing a new calculus (Itoh). One may characterize the approaches glibly as the physicist's (S) or the mathematician's (I). We shall adopt the former should the need arise.

Enter, Temperature

Now let the medium in which diffusion takes place be endowed with a temperature, whereupon we expect the Brownian particle to come into equilibrium with its host in the sense that $<V(t)^2> \to \frac{k_B T}{M}$ after a sufficiently long time. The result is quite independent of initial conditions. Since $<V_1(t)^2> = \frac{\gamma}{\beta}(1-e^{-2\beta t})$ and $<V_1(t)^2> \to <V(t)^2>$, we conclude that $\frac{\gamma}{\beta} = \frac{k_B T}{M}$. This relation is another example of a *Fluctuation-Dissipation Theorem*—the fluctuation, γ, being related to the dissipation, β, via $k_B T$, a quantity associated with thermodynamic equilibrium. The various probability
distributions discussed earlier may be modified to include the approach to equilibrium by replacing $\frac{\gamma}{\beta}$ by $\frac{k_B T}{M}$. One sees, then, that

$$P(v,t|v_0,0) \to \frac{1}{[2\pi\frac{k_B T}{M}]^{\frac{1}{2}}} \exp[-\frac{Mv^2}{2k_B T}] \ , \tag{4.19}$$

the Maxwell distribution, as expected. And the distribution satisfies the equation

$$\frac{\partial P(v,t)}{\partial t} = D_v \frac{\partial}{\partial v}(\frac{\partial}{\partial v} + \frac{M}{k_B T} v)\, P(v,t) \tag{4.20}$$

with $P(v,t=0)=\delta(v-v_0)$. The equation bears the names of the physicists A. D. Fokker and Max Planck. The quantity $D_v=\beta\frac{k_BT}{M}=\gamma$ may be regarded as a diffusion co-efficient in velocity. It is a measure of rapid, small changes in velocity brought about by the stochastic force.

Another example of the role played by the thermal host occurs when, in statistical mechanics, we find it useful to average over initial states.

Consider the variance of $X_1(t)$, $\sigma_x^2(t)=\int_0^t dt'\int_0^t dt''<V_1(t')V_1(t'')>$. The integrand is interesting in its own light. It is symmetric, and

$$<V_1(t')V_1(t'')>=\int_0^{t'}dt_1\int_0^{t''}dt_2\,e^{-\beta(t'-t_1)}e^{-\beta(t''-t_2)}<L(t_1)L(t_2)> \qquad (4.21)$$

$$=\frac{\gamma}{\beta}[e^{-\beta|t'-t''|}-e^{-\beta(t'+t'')}].$$

Expanding the lhs brings us

$$<V(t')V(t'')>=\frac{\gamma}{\beta}e^{-\beta|t'-t''|}+[v_0^2-\frac{\gamma}{\beta}]\,e^{-\beta(t'+t'')}. \qquad (4.22)$$

Again, note that while the fluctuations, e.g. $<V_1(t')V_1(t'')>$, are independent of the initial state, correlations between dynamical quantities, $<V(t')V(t'')>$, most certainly are not.

As before, we set the harassed particle in a fluid which is at equilibrium at temperature T, and note that i) the particle will, in time, come into thermal equilibrium with the fluid, and ii) we propose that initially, the particle is chosen from an equilibrium ensemble, with respect to which we will average. Since the thermal average $\{\ldots\}$ of v_0^2 is $\frac{k_BT}{M}$, we obtain, after the "double averaging,"

$$\{<V(t')V(t'')>\}=\frac{k_BT}{M}\,e^{-\beta|t'-t''|}\equiv C_V(t',t''). \qquad (4.23)$$

This correlation function for velocity is symmetric in (*t', t''*) and vanishes, slowly and exponentially. The *stationary and homogeneous* character of the stochastic process—its underlying randomness, its setting, does not vary with time This implies that the function will depend only upon the difference in times, (*t'-t''*).

Now consider the doubly-averaged $\{<X^2(t)>\}$, a measure of the displacement

$$\{< X^2(t) >\} = \int_0^t dt' \int_0^t dt'' \{< V(t')V(t'') >\}. \tag{4.24}$$

Since the correlation function depends only upon the absolute difference of its arguments, we have, after integration, with $C_V(t) = \{< V(t)V(0) >\}$,

$$\{< X^2(t) >\} = 2\int_0^t dt'(t - t')C_V(t'). \tag{4.25}$$

At long time intervals, provided that the decay of the correlation is rapid enough, we have

$$\{< X^2(t) >\} \rightarrow \; 2t \int_0^\infty dt' \; C_V(t') = 2Dt \; . \tag{4.26}$$

But this is precisely the behavior expected from simple diffusion. The representation of a transport coefficient—in this case the diffusion co-efficient as an integral of a correlation function—is sometimes called a *Kubo relation.* In our particular case integration gives $D = \frac{2k_B T}{M\beta}$. There is some interesting physics here: typically, a diffusion coefficient describes a "jump," or *random walk* process, which contains both a jump, or flight-length, and a frequency of jumping—or a mean velocity for the random walker. These features may be noted here, for $\frac{2k_B T}{M}$ is the square of a thermal velocity, say v_T^2, and $\frac{v_T}{\beta}$ is the distance traveled following an initial impulsive velocity, before damping destroys inertia.

Lecture 5

Stochastic Processes

When we wish to consider the temporal evolution of a random quantity, perhaps a diffusing molecule, a chemical reaction, or an evolving population, we are led to the notion of *stochastic process.* This branch of the subject deals with variables, e.g., $X(t)$, whose distribution of outcomes, $p(x,t)$, varies with time. In a sense we deal with an infinity of stochastic variables, labeled by *t*. If we consider measurements carried out at $t_1, t_2, \ldots t_N$ we are led to inquire about $p_N(x_1,t_1;\ x_2,t_2;\ldots;x_N,t_N)$, the probability (density) for the variable to assume value x_1 at t_1, x_2 at t_2, etc. Hence the notion of "process." But there is another, quite rich, point of view. In a homely analogy, the arguments of the density p_N remind one of a train schedule, station x_1 at t_1, station x_2 at t_2, etc. We have an *N*-point sampling of a trajectory, a function (not necessarily smooth!), $x = F(x_0,t)$. And p_N is a number associated with the choice. This particular trajectory is one of a collection, an *ensemble*, of functions; here we have chosen a particular *representative,* or *realization* of the process. Finally, we will average over all members of the ensemble.

For a more formal treatment, let there be a stochastic variable X assuming values *x* with associated distribution $P(x)$. Now let us form $F(X,t) \equiv X(t)$ a function of X and parameter t such that $F(X,0) \equiv X \equiv X(0)$. The function assumes values $F(x,t)$. Then,

$$p_1(x,t) = \int dx_0\ P(x_0)\,\delta(x - F(x_0,t)), \tag{5.1}$$

and we can construct an *N*-time distribution for $X(t)$ as

$$p_N(x_1,t_1;x_2,t_2;\ldots x_N,t_N) = \int dx_0\ P(x_0)\ \delta(x_1 - F(x_0,t_1))\ \delta(x_2 - F(x_0,t_2))\ldots\delta(x_N - F(x_0,t_N)) \tag{5.2}$$

which is symmetric upon interchange of (*x*, *t*) pairs. The process is defined by the initial distribution $P(x)$ and the function $F(X,t)$. It makes the "train schedule" image precise: we select a trajectory $F(x_0,t)$ by choosing the initial value x_0. We choose it with probability (density) $P(x_0)$. Then, if that trajectory passes through x_1 at t_1, x_2 at t_2, etc., the integrand is non-zero, making a contribution to p_N. In most cases of interest the ensemble of trajectories is an ensemble of very irregular functions.

Some consequences of Eq. (5.2) are

$$< X(t_1)\, X(t_2) > = \int dx_1 ... \int dx_N \;\; p_N(x_1,t_1\,;\, x_2,t_2;... x_N\, t_N)\; x_1 x_2$$
$$= \int dx_1 \int dx_2 \;\; p_2(x_1,t_1;x_2,t_2)\; x_1\, x_2 \;=\; \int dx_0\; P(x_0)\; F(x_0,t_1)\; F(x_0,t_2)\;, \tag{5.3}$$

with obvious generalization to $< X(t_1)\, X(t_2) X(t_3) >$ and correlations of higher order. And we shall use "condensing" relations of the sort

$$\int dx_3 \int \; dx_4 ... \int dx_N \;\; p_N(x_1,t_1\,;\, x_2,t_2;... x_N\, t_N) \;=\; p_2(x_1,t_1\,;\, x_2,t_2) \tag{5.4}$$

to produce marginal, or contracted, distributions. (Note that the associated time-variables vanish when the *x*-integrations are performed.)

Conditional probabilities play an important role. As examples,

$$p_3(x_1,t_1\,;\, x_2,t_2;\, x_3,t_3) = p(x_2,t_2\,;\, x_3,t_3 \mid x_1,t_1)\; p_1(x_1,t_1) = p(x_3,t_3 \mid x_1,t_1\,;\, x_2,t_2)\; p_2(x_1,t_1;x_2,t_2)$$

have a simple interpretation: One might ask, "What evolution is consequent to my beginning at (x_1,t_1)?" or, "Having experienced $(x_1,t_1\,;\, x_2,t_2)$, what can I expect at t_3?" Finally, should we be dealing with a situation in which the randomness is *stationary*, the various probabilities will be unchanged if all time arguments are displaced by a common amount.

Markov Processes

Stochastic processes are complicated. A considerable simplification occurs if we make the assumption that

$$p(x_N,t_N \mid x_1,t_1;\, x_2,t_2;...;x_{N-1},t_{N-1}) \;=\; p(x_N,t_N \mid x_{N-1},t_{N-1})\,, \tag{5.5}$$

which asserts that after a long journey, the trajectory having passed through 1, 2, 3 …, the probability of x_N at t_N depends only upon the most recent pair (measurement) (x_{N-1},t_{N-1}). Such an assumption implies a considerable loss of memory. Nevertheless, almost all treatments of stochastic processes in the sciences appear to be based upon this approximation, sometimes justified, sometimes questionable. If we condense our notation, writing $(x_N,t_N) \to N$, the Markov approximation may be written

$$p(N \mid 1,2,...N-1) \;=\; p(N \mid N-1)\,.$$

In particular,

$$p_2(1,2) = p(2|1)\, p_1(1),$$

$$p_3(1,2,3) = p(3|2)\; p(2|1)\; p_1(1) \tag{5.6}$$

$$\cdots$$

$$p_N(1,2,...N) = p(N|N-1)\, p(N-1|N-2)...p(2|1)\, p_1(1).$$

All is determined by a single function, $p(2|1) = p(x_2, t_2 | x_1, t_1)$. If the statistics are *stationary* the function will depend only upon the difference between times. Then we have the important *transition probability.*

$$p(x_2, t_2 | x_1, t_1) \equiv T(x_2 | x_1; \tau) \qquad \tau = t_2 - t_1.$$

One example of a transition probability is:

$$P(v,t|v(0),0) = \frac{1}{[2\pi \frac{k_B T}{M}(1-e^{-2\beta t})]^{\frac{1}{2}}} \exp[-\frac{M}{2k_B T}\frac{(v - v(0)e^{-\beta t})^2}{(1-e^{-2\beta t})}] \tag{5.7}$$

which appeared in our study of the *Langevin Equation*. It is the transition probability describing the *Ornstein-Uhlenbeck* process. Another is

$$\overline{P}(x|x_0;\tau) = \frac{1}{\sqrt{2\pi\tau}} \exp[-\frac{(x-x_0)^2}{2\tau}]\ , \tag{5.8}$$

the transition probability for the diffusive, or *Wiener-Levy* process, which we encountered in our study of the random walk.

An immediate and important consequence of the Markov assumption proceeds from a comparison. On the one hand

$$\int dx_2\ p_3(1,2,3) = p_2(1,3) = p(3|1)\, p_1(1) \tag{5.9}$$

is generally true. (The result is independent of the choice of t_2.) But for the Markov process we have

$$\int dx_2\ p(3|2)\, p(2|1)\, p_1(1) = p(3|1)\, p_1(1),$$

whence

$$\int dx_2\ p(x_3, t_3 | x_2, t_2)\, p(x_2, t_2 | x_1, t_1) = p(x_3, t_3 | x_1, t_1). \tag{5.10}$$

This is the Chapman-Kolmogorov equation, a result of great importance to our development of the subject. (The form is reminiscent of a matrix-multiplication.) If we multiply both sides by $p_1(x_1,t_1)$—to introduce p_2—then integrate with respect to x_1, we retrieve an alternate form,

$$\int dx_2\ p(x_3,t_3\,|\,x_2,t_2)\,p_1(x_2,t_2) = p_1(x_3,t_3)\,. \tag{5.11}$$

In the *stationary* case,

$$\int dx_2\,T(x_3\,|\,x_2;\,t_3-t_2)\,T(x_2\,|\,x_1;\,t_2-t_1) = T(x_3\,|\,x_1;t_3-t_1)\,. \tag{5.12}$$

Master Equations

The Chapman-Kolmogorov equation may be converted to an integro-differential equation in the *stationary* case, by first writing $t_3-t_2=\tau$, $t_2-t_1=t$, $t_3-t_1=t+\tau$, whence

$$T(x_3\,|\,x_1;t+\tau) = \int dx_2\,T(x_3\,|\,x_2\,;\tau)\;T(x_2\,|\,x_1;t)\,. \tag{5.14}$$

Now we let $\tau\to 0$, noting that $T(x_3\,|\,x_2\,;\tau)\to\delta(x_3-x_2)+\tau\,K(x_3,x_2)$ + "smaller" and $\int dx_3\ K(x_3,x_2) = 0$ (normalization). We obtain

$$\frac{\partial T(x_3\,|\,x_1;t)}{\partial t} = \int dx_2\ K(x_3,x_2)\ T(x_2\,|\,x_1;t)\,. \tag{5.15}$$

It is convenient to write $K(x_3,x_2) = W(x_3,x_2)-\delta(x_3-x_2)\int dx_3\,W(x_3,x_2)$, which preserves the normalization. Then,

$$\frac{\partial T(x_3\,|\,x_1;t)}{\partial t} = \int dx_2\ W(x_3,x_2)\ T(x_2\,|\,x_1;t) - \{\int dx_2\,W(x_2,x_3)\}\,T(x_3\,|\,x_1;t)\,. \tag{5.16}$$

(Overall, we have set

$$T(x_3\,|\,x_2\,;\tau)\to\delta(x_3-x_2)[1-\tau\int dx_3\,W(x_3,x_2)]+\tau\,W(x_3,x_2)\ +\text{"smaller".)}$$

Usually, we are most interested in $p_1(x_2,t)$, the distribution evolving from $p_1(x_2,0)$. Since

$$p_1(x_2,t) = \int dx_1\,T(x_2\,|\,x_1;t)\,p_1(x_1,0)\,, \tag{5.17}$$

we may obtain an equation for $p_1(x_2,t)$ by multiplying the equation for T(.|.) by $p_1(x_1,0)$ and integrating. We obtain a master equation in its usual form,

$$\frac{\partial p_1(x_1;t)}{\partial t} = \int dx_2\, W(x_1, x_2)\, p_1(x_2;t) - \{\int dx_2\, W(x_2, x_1)\}\; p_1(x_1;t)\,. \quad (5.18)$$

The equation may be described as a statement of balance, or bookkeeping. The first term on the rhs describes processes feeding x_1 from all x_2, while the second describes all processes leaving x_1 for the various x_2. It is consistent with this view—and dimensionally correct—to represent W as $W(x_1, x_2) = \pi(x_1 \mid x_2)\, w(x_2)$, where $\pi(x_1 \mid x_2)$ is the (conditional) probability for $x_2 \to x_1$ and $w(x_2)$ is a "rate" characterizing the process. Thus

$$\int dx_1\, \pi(x_1 \mid x_2) = 1 \qquad \text{and} \qquad \int dx_2\, W(x_2, x_1) = w(x_1)\,, \quad (5.19)$$

and the master equation may be written

$$\frac{\partial p_1(x_1;t)}{\partial t} + w(x_1)\, p_1(x_1;t) = \int dx_2\, \pi(x_1 \mid x_2)\, w(x_2)\, p_1(x_2;t)\,. \quad (5.20)$$

We shall now enrich the equation by including deterministic effects along with the stochastics. To begin simply, consider the deterministic version, or limit, of $T(x_3 \mid x_2;\tau)$. Suppose we were describing the uninhibited motion of a single particle along a line. Its speed v is constant, its initial position, stochastic. Then,

$$T(x_3 \mid x_2;\tau) = \delta(x_3 - x_2 - v\tau) \to \delta(x_3 - x_2) - \tau v\, \delta'(x_3 - x_2) + o(\tau)$$

$$\text{and using } K(x_3, x_2) = -v\, \delta'(x_3 - x_2)$$

in Eq. (5.15) brings us to

$$\frac{\partial T(x_3 \mid x_1;t)}{\partial t} + v \frac{\partial T(x_3 \mid x_1;t)}{\partial x_3} = 0\,.$$

Should we be dealing with motion in the phase space *(x,v)*, the particle being accelerated by an external field of force, *F(x)*, the generalization would be

$$\begin{aligned} T(x_3, v_3 \mid x_2, v_2;\tau) &= \delta(x_3 - x_2 - v\tau)\, \delta(v_3 - v_2 - a(x_2)\tau) \\ &= \delta(x_3 - x_2)\, \delta(v_3 - v_2) - \tau \{v_2\, \delta'(x_3 - x_2)\, \delta(v_3 - v_2) + a(x_2)\, \delta(x_3 - x_2)\, \delta'(v_3 - v_2)\} \\ &\quad + o(\tau) \end{aligned}$$

and

$$\frac{\partial T(x_3,v_3\,|x_1,v_1;t)}{\partial t} + \; v_3\frac{\partial T(x_3,v_3\,|x_1,v_1;t)}{\partial x_3} + a(x_3)\frac{\partial T(x_3,v_3\,|x_1,v_1;t)}{\partial v_3} = 0\,, \tag{5.21}$$

where *a(x) = F(x)/m*. The next step, which we will describe in the following lecture, is to combine the deterministic and stochastic processes.

In another preview, note that when we deal with discrete states we will have

$$\frac{\partial\, p(n;t)}{\partial t} = \; \sum_{\ell} W(n,\ell)\, p(\ell;t) - \{\sum_{\ell} W(\ell,n)\}\; p(n;t)\,, \tag{5.22}$$

whose further refinement will be dealt with later. What began as an almost trivial relationship, "*T(3,1)=T(3,2)* T(2,1)*," has now become a rich and imposing equation. The several equations we have been discussing—in their various versions—1) continuous time/continuous range of states, 2) continuous time/discrete states, 3) discrete time /discrete states (i.e., jumps and Random Walk!) are used throughout physics, chemistry, biophysics and engineering. When, in the discrete case, the number of states is finite, we often refer to the system as a *Markov chain.*

There is something rather strange about what we have been developing, for we have done little more than manipulate a formal statement about the transition probability. Where's the physics? Consider two points: first, that master equations generate the full transition probability from the mere knowledge of its short-term behavior; and second, that the short-term behavior, the kernel $W(x_1\,,x_2)$ contains the physics. The kernel might describe scattering processes, classical or quantum-mechanical. Quite generally, it may be extracted from Fermi's Golden Rule, the famous formula describing quantum - mechanical perturbation theory.

As a final comment, we might note that a small modification of our treatment of the Chapman –Kolmogorov equation leads to

$$p(x_3,\, t+\tau) = \int dx_2\, T(x_3\,|x_2\,;\tau)\;\, p(x_2,t)\,,$$

an equation first proposed and used by Albert Einstein, in his analysis of Brownian motion. (Here, *x* refers to the position of a particle.) The success of that work helped him to his Nobel Prize in physics.

Lecture 6

Markov Processes: Applications and Examples

We continue with the case of "continuous time, continuous variable." This is the model that pervades kinetic theory:

$$\frac{\partial p_1(x_1;t)}{\partial t} = \int dx_2\, W(x_1, x_2)\, p_1(x_2;t) - \{\int dx_2\, W(x_2, x_1)\}\; p_1(x_1;t) \qquad (6.1a)$$

or

$$\frac{\partial p_1(x_1;t)}{\partial t} + w(x_1)\, p_1(x_1;t) = \int dx_2\, \pi(x_1 \mid x_2)\, w(x_2)\, p_1(x_2;t)\,. \qquad (6.1b)$$

To give an example of the usefulness of Eq. (6.1b), let x represent v, the 3-dimensional velocity of a particle (impurity, ion, neutron, photon, etc.) diffusing through a host material. Multiply the equation by N, so that $N\; p_1(x_1;t) \to n(v,t)$. Then, the equation describes the evolution/relaxation of the distribution of particle velocities through a series of collisions with targets comprising the host material. (We assume that the particles are distributed uniformly throughout space.) The quantity $w(v)$, the "rate," is compounded of the density of scatterers, the speed of the diffusing particle, and the cross-section for scattering. If the host is characterized by a temperature, a bit of physics beyond the kinematics of the collision must be introduced. It is the notion of *detailed balance.* We must have

$$W(v_1, v_2)\, M(v_2, T) = W(v_2, v_1)\, M(v_1, T)\,, \qquad (6.2)$$

where $M(v_1, T)$ is the Maxwell-Boltzmann distribution appropriate to temperature T. Detailed balance helps insure that the particles will, after a sufficiently long time, become thermalized.

We may generalize the foregoing example easily, to include the effects of streaming and of acceleration by a force field. As in the last lecture, the symbol x will represent both position and velocity. The transition probability will describe deterministic as well as stochastic effects (collisions.) The key bit of physics for the distinction is that the fundamental time interval is assumed to be long compared the duration of a single collision, but short compared with the time between collisions. The transition probability will then be the product of a factor describing the deterministic motion and another describing the stochastic (collisions). Expansion in this suitably small time interval yields a transition probability in which the two components simply add. When we grant the argument and review the calculation presented at the end of the last lecture, we are led easily to the generalization

$$\frac{\partial p_1(x_1,v_1;t)}{\partial t} + v_1 \cdot \frac{\partial p_1(x_1,v_1;t)}{\partial x_1} + a(x_1) \cdot \frac{\partial p_1(x_1,v_1;t)}{\partial v_1} =$$
$$\int dv_2\, \pi(v_1 \mid v_2)\, w(v_2)\, p_1(x_1,v_2;t) - w(v_1)\, p_1(x_1,v_2;t), \qquad (6.3)$$

a full-blown transport equation. When multiplied by *N*, the equation describes the transport of *N* independent particles. Note that the solution to this equation is a probability distribution, containing information about higher moments and fluctuations. Most of the time, however, these features are neglected, and attention is placed upon expected values of some dynamical variable, *f(x)*. Thus, $\int dx_1\ p_1(x_1)\, f(x_1) \equiv < f >$. These mean values may also be described by the equation. As an example, let $N\,\delta(x_1 - x)$ be the "*f* " describing particle density at *x*. If we multiply the equation for $p_1(x_1,t)$ by *f* and integrate, we obtain an equation for the expected value of the density at *x*, an equation that appears to be the same as the original. Only the labeling (and meaning!) are different.

The Fokker-Planck Equation

Typically, master equations are integro-differential equations and, though linear, are not amenable to analytic solution. But when the jumps are in some sense small, a master equation can be simplified. It becomes a partial differential equation of second order, the Fokker-Planck equation. This equation is relevant, for example, when collisions occur between charged particles and when light diffuses through complex media. In both cases the angular deviation associated with a single scattering is small. The literature contains many derivations of the Fokker-Planck equation; one is sketched in the Appendix.

The Fokker-Planck equation has the form of a drift-diffusion equation

$$\frac{\partial P(x,t)}{\partial t} + \frac{\partial}{\partial x} D_1(x) P(x,t) = \frac{1}{2} \frac{\partial^2}{\partial x^2} D_2(x) P(x,t), \qquad (6.4)$$

and is easier to deal with, analytically, than is the integro-differential master equation.

When the equation describes transport processes, $D_1(x) > 0$ is called the drift coefficient and $D_2(x) > 0$ the diffusion coefficient. These coefficients are related to moments of the transition probability, in that

$$T_n(x;\tau) = \int dx_1\, (x_1 - x)^n\, T(x_1, x\ ;\tau) = D_n(x)\,\tau \qquad \text{for } n = 1, 2\ .$$

We have met differential equations of this form before. The simple diffusive process gives

$$\frac{\partial P(x,t)}{\partial t} = D\,\frac{\partial^2}{\partial x^2} P(x,t)\,,$$

and Brownian motion, the Ornstein-Uhlenbeck process, gives

$$\frac{\partial P(v,t)}{\partial t} = \gamma \frac{\partial}{\partial v}\left(\frac{\partial}{\partial v} + \frac{M}{k_B T} v\right) P(v,t), \tag{6.5}$$

which is Fokker-Planck with a constant diffusion coefficient and a linear drift, $\beta v = (\gamma M / k_B T) v$. Crucial to the derivation from the integro-differential equation is the requirement that moments of the transition probability higher than the second are vanishingly small in the limit of small jump size. We can verify that this property prevails in the two transition probabilities we have encountered. For example,

$$\overline{P}(x,\tau) = \frac{1}{\sqrt{4\pi D \tau}} \exp[-\frac{x^2}{4D\tau}] = T(x,0;\tau)$$

describes the diffusion process, whereupon

$$\int_{-\infty}^{\infty} dx\, x^n\, T(x,0;\tau) = \frac{1}{\sqrt{4\pi D t}} \int_{-\infty}^{\infty} dx\, x^n \exp[-\frac{x^2}{4D\tau}]$$

vanishes for odd integers *n* and is proportional to τ^m when *n=2m* is even. A similar result follows easily from the Ornstein-Uhlenbeck process. It is not at all obvious that most physical processes will possess this property.

We may enrich the Fokker-Planck equation to include deterministic evolution as well, in the manner used earlier. Then, we are led to the rather general kinetic equation, of form

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x} + \frac{F(x)}{M}\frac{\partial}{\partial v}\right) P(x,v\,|\,..,t) = \left(-\frac{\partial}{\partial v} D_1(v) + \frac{1}{2}\frac{\partial^2}{\partial v^2} D_2(v)\right) P(x,v\,|\,..,t)\,. \tag{6.6}$$

Most applications use the Ornstein-Uhlenbeck process so that

$$\left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x} + \frac{F(x)}{M}\frac{\partial}{\partial v}\right) P(x,v\,|\,..,t) = \beta\left(\frac{\partial}{\partial v}\, v + \frac{k_B T}{M}\frac{\partial^2}{\partial v^2}\right) P(x,v\,|\,..,t)\,, \tag{6.7}$$

an equation considered by H. A. Kramers and O. Klein. Note that just as the space-independent equation supported the stationary solution

$$P(v) = C\,Exp[-\frac{M\,v^2}{2k_BT}],$$

so the more general equation supports

$$P(x,v) = C\,Exp[-\frac{1}{k_BT}(\frac{1}{2}M\,v^2 + V(x))].$$

Beyond Ornstein-Uhlenbeck

As long as the moments of the transition probability have the Fokker-Planck property, the master equation will be a linear, partial differential equation of second order. Indeed, it may be regarded as an equation of Schroedinger type, and associated with it are many techniques of approximation. One assumes that such an equation is easier to deal with than an integro-differential equation. (But in these days of easy access to powerful computers . . . .)

The Fokker-Planck equation will generate the transition probability, with its full time-dependence, once its two transfer moments are known. And these are determined by the short-time behavior of the system (Markov again!) extracted from an underlying physical model. For one class of models, whose physics can be described by a generalization of the Langevin equation, we can proceed easily. Let

$$\frac{dx(t)}{dt} = f(x) + g(x)\,L(t) \tag{6.8}$$

with $L(t)$ the white noise we have encountered in the past, $<L(t)>=0$, $< L(t)\,L(t') >= 2\,\gamma\,\delta(t-t')$. If there is a problem in interpreting the stochastic driver, we will take the Stratonovich view.

To get the transfer moments, we write

$$\Delta x(\tau) \equiv x(\tau) - x_0 = \int_0^\tau dt\ [\ f(x(t)) + g(x(t))\,L(t)]$$

and expand to (6.9)

$$\Delta x(\tau) = f_0\,\tau + \int_0^\tau dt\ [\,g_0 + g'_0\ \Delta x(t)\,]L(t) + o(\tau),$$

with $x(0) = x_0$, $f(x_0) = f_0$, $g(x_0) = g_0$, $g'(x_0) = g'_0$.

We will regard x_0 as fixed, $x(\tau)$ as stochastic. When we average, we find

$$< \Delta x(\tau) >= f_0\,\tau + g'_0 \int_0^\tau dt\ \ < \Delta x(t)\ L(t) > + o(\tau).$$

The integrand is evaluated by iteration of the integral equation. To leading order we have

$$< \Delta x(\tau) >= f_0\, \tau + g'_0\, g_0 \int_0^\tau dt \int_0^t dt' \; < L(t)\; L(t') > + o(\tau),$$

from which we infer $< \Delta x(\tau) >= (f_0 + 2\gamma\, g'_0\, g_0)\tau$. Since the $< (\Delta x(\tau))^n >$ are the moments of the transition probability, we may infer a drift co-efficient

$$D_1(x) = f(x) + 2\gamma\, g'(x)\, g(x). \tag{6.10}$$

The latter contribution is sometimes called "noise-induced" drift.

The diffusion coefficient stems from $< (\Delta x(\tau))^2 >$. If we contemplate averaging the square of the equation for $< \Delta x(\tau) >$ we see that the only term contributing to $O(\tau)$ is $g_0^2 \int_0^\tau dt \int_0^\tau dt' \; < L(t)\; L(t') > = 2\gamma\, g_0^2\, \tau$, which yields

$$D_2(x) = 2\gamma\, g^2(x). \tag{6.11}$$

.A similar analysis may be performed on an *n*-variable system.

Strong Damping-Smoluchowski's Equation

As a final example of master equations which occur in the form of differential equations, consider the Langevin system

$$\frac{d^2x}{dt^2} + \beta \frac{dx}{dt} = \frac{1}{M} F(x) + L(t) \tag{6.12}$$

and make the assumption that the motion is heavily damped. Thus, the inertial term, the second derivative, is negligible. (This motion is characteristic of biological systems.) The equation that results is a special case of the classic Langevin equation. (Imagine $v \to x$ in that model.) There is, formally, no damping; the process is purely diffusive in the presence of an external field. The equation is now of form (6.8) with $f(x) = \frac{1}{\beta M} F(x)$ and $g(x) = \frac{1}{\beta}$. If, further, to ensure thermal equilibrium, we require that $\frac{\gamma}{\beta} = \frac{k_B T}{M}$, we are led to

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} \frac{F(x)}{M\beta} \right) P(x,t) = \frac{k_B T}{M\beta} \frac{\partial^2}{\partial x^2} P(x,t) \qquad D = \frac{k_B T}{M\beta} \tag{6.13}$$

to describe diffusion under the influence of an external field and strong damping. *D* is the usual coefficient of diffusion. This useful equation is named after the Polish physicist Marian Smoluchowski. Note that the stationary solution is

$$P(x) = C\,Exp[-\frac{V(x)}{k_B T}\ ],$$

and suggests a situation in which inertial effects may be neglected. Another way of viewing this "large β" approximation is to assert that we are "averaging" or "coarse-graining" the full dynamics over time intervals of order $(1/\beta)$.

Lecture 7

Markov Processes with Discrete States

We turn to the class of systems described by discrete states and continuous time. Applications here might be to studies of transitions between the discrete quantum states of a system or, in a leap across boundaries, to population studies in biology or environmental science.

Let us return to the Chapman-Kolmogorov equation in the case of discrete states. In particular, we study a variable *X(t)* whose values are restricted to the non-negative integers. We will alter notation somewhat, to write the equation as

$$\sum_{m=0} P(n,t_3 \mid m,t_2)\,P(m,t_2 \mid \ell,t_1) = P(n,t_3 \mid \ell,t_1) \qquad t_3 \geq t_2 \geq t_1 . \tag{7.1}$$

Thus, we have transitions from ℓ at t_1 to n at t_3 via an intermediate state *m* at an arbitrary time t_2. We convert the expression to a differential equation by examining the behavior of the conditional probabilities when the time differences are small.

Consider $P(n,t+dt \mid m,t)$, beginning with the case *n = m.* Then

$$P(m,t+dt \mid m,t) = 1\text{- } a(m,t)\,dt\ , \tag{7.2}$$

with *a(m,t) dt* the probability that transition from state *m* occurs during the time interval *dt*. We may write

$$P(n,t+dt \mid m,t) = a(m,t)\,p(n,m;t)\,dt\ , \qquad n \neq m \tag{7.3}$$

with the new conditional probabilities *p(n,m;t)* (the probability that <u>if</u> something happens it will be a transition to state *n*). Of course $\sum_n p(n,m;t) = 1$. (We take *p(n,n;t)* = 0.)

If these expressions are used in the (C-K) equation for $P(n,t_3 \mid \ell,t_1)$ with $t_3 = t_2 + dt$, and if one keeps track of indices in the sum, one finds, in the usual manner, the famous *forward equations,*

$$\frac{\partial}{\partial t_2} P(n,t_2 \mid \ell,t_1) + a(n,t_2) P(n,t_2 \mid \ell,t_1) = \sum_m a(m,t_2)\,p(n,m;t_2)\,P(m,t_2 \mid \ell,t_1). \tag{7.4}$$

Each is a discrete version of the master equation featured in Lecture 6. The normalization, $\sum_n P(n,t_2 \mid \ell,t_1) = 1$, is easy to verify.

A set of *backward equations*,

$$\frac{\partial}{\partial t_1} P(n,t_2 \mid \ell,t_1) - a(n,t_1) P(n,t_2 \mid \ell,t_1) = -\, a(\ell,t_1) \sum_m P(n,t_2 \mid m,t_1)\, p(m,\ell;t_1), \tag{7.5}$$

may be generated in a similar manner, but they will not concern us. Both sets of equations must be augmented by proper initial conditions. When we "model" a process we choose appropriate functions *a(n,t)* and *p(n,m;t).* The choice is familiar to practitioners of particle transport theory. It is the choice of both total cross-section and the partial (differential) cross-sections which sum to the total.

Now, several simplifications: First, the symbol *n*, which generally describes a state, in this case indexes a population of *n* statistically independent objects. Then, the function $a(n,t)$, which is proportional to the probability of "something happening" to this population, will usually be of the form *n a(t)* . Second, if we deal with *stationary* randomness, the underlying milieu being independent of time, the functions $a(n,t)$ and *p(n,m; t)* will be independent of time. Functions bearing two times will depend only upon the time difference. Thus,

$$\frac{\partial}{\partial t} P(n \mid \ell;t) + a(n)\, P(n \mid \ell;t) = \sum_m a(m)\, p(n,m)\, P(m \mid \ell;t). \tag{7.6}$$

(Recall that $P(n \mid \ell;t) \equiv T(n \mid \ell;t)$.)

We infer from this equation for the conditional probabilities that the probability density satisfies

$$\frac{\partial}{\partial t} p_1(n,t) + a(n)\, p_1(n,t) = \sum_m a(m)\, p(n,m)\, p_1(m,t)\ , \tag{7.7}$$

with $p_1(n, t=0)$ prescribed. That the "conservation of probability" prevails may be seen by summing the equation(s) with respect to *n.* And, for many models, *a(n)= n a*.

Birth and Death Processes

In many situations the process in question involves populations subject to births and deaths. Then, while upward transitions (births) from *m* to *n* are relatively un-hindered, downward transitions (deaths) from *m* to *n* are limited to *m-n*=1 .The process is usually *homogeneous*, too, in that the probability *p(n,m)* depends upon (*n,m*) only through *(n-m)*. Then we may write *p(n,m) = f(n*+1*-m)* and

$$\frac{1}{a}\frac{\partial}{\partial t} P(n \mid \ell;\, t) + n\, P(n \mid \ell;t) = \sum_{m=0}^{n+1} m\, f(n+1-m)\, P(m \mid \ell;\, t). \tag{7.8}$$

These equations may be solved by considering the associated generating function, $G(z \mid \ell; t)=\sum_{n=0}^{\infty} z^n P(n \mid \ell; t)$. If we multiply the equation by z^n and sum, recalling that

$$\sum_{n=0}^{\infty} n z^n P(n \mid \ell; t)=z\frac{\partial}{\partial z}G(z \mid \ell;t) \quad \text{and} \quad \sum_{n=0}^{\infty} z^n \sum_{m=0}^{n} p_1(n-m)p_2(m)= g_1(z)\,g_2(z),$$

where the *g's* are the corresponding generating functions, we are led to the quite general expression

$$\frac{1}{a}\frac{\partial}{\partial t}G(z \mid \ell; t) + [z-g(z)]\frac{\partial}{\partial z}G(z \mid \ell; t) = 0, \tag{7.9}$$

with $g(z)=\sum_{n=0}^{\infty} z^n f(n)$. In one set of examples, we describe a simple birth β, death γ, process (one offspring—no twins, triplets, …) ($\gamma+\beta=1$). Their respective probabilities are

$$p(n,m)=\gamma\,\delta(n-m+1)+\beta\,\delta(n-m-1), \qquad f(\ell)=\gamma\,\delta(\ell)+\beta\delta(\ell-2).$$

and we obtain

$$[g(z)-z]=(z-1)(\beta z-\gamma). \tag{7.10}$$

Examples

We choose a simple birth process for the first example. Here, we return to the beginning of our treatment and take

$$a(m,t)=\beta, \qquad p(n,m;t)=\delta(n-m-1).$$

Then the probability of transition between the allowed states during interval *dt* is simply $\beta\,dt$. (Note that in this model, the likelihood of birth is independent of the existing population. The birth event may also be thought of as the arrival of a signal.)

We will be generating a Poisson process, a Poisson time-series. The forward equations are

$$\frac{1}{\beta}\frac{\partial}{\partial t}P(n \mid \ell;t)+P(n \mid \ell;t)=P(n-1 \mid \ell;t) \qquad P(n \mid \ell;t)=0,\ n<\ell\ .$$

Then,

$$\frac{\partial G(z\,|\,\ell;\,t)}{\partial t}+\beta(1-z)\,G(z\,|\,\ell;\,t)=0\,, \tag{7.11}$$

whose solution, for $P(n\,|\,\ell,0)=\delta(n,\ell)$, is,

(7.12a)

$$G(z\,|\,\ell;\,t)=z^{\ell}e^{-(1-z)\beta t}\Rightarrow P(n\,|\,\ell,t)=e^{-\beta t}\,\frac{(\beta t)^{n-\ell}}{(n-\ell)!}\quad n\geq\ell$$

—the Poisson distribution, with the mean number of particles increasing as βt.

These results can be used in a variety of ways. For example, the "births" might describe proteins, sired by a gene, and sent on a mission. Or, the events might be the turning on or off of a switch. Finally, they might represent the arrival of electrons at a detector, in a situation in which there is no correlation between successive arrivals, the probability of arrival being simply proportional to *dt*. The statistics of the current which ensues is Poisson, and the fluctuations are related to *shot noise*.

The companion to this model might be a simple death-process, in which the probability of loss is independent of the size of the collection. It may depend solely upon the ability of the "sink" to process. (Imagine a single clerk disposing of a collection of complaints.) Now we have

$$a(m,t)=\gamma\,,\quad p(n,m;t)=\delta(n-m+1)$$

$$\frac{1}{\gamma}\frac{\partial}{\partial t}P(n|\,\ell;t)+P(n\,|\,\ell;t)=P(n+1|\,\ell;t)\qquad P(n\,|\,\ell;t)=0\,,\ n>\ell\,.$$

These equations do not lend themselves to solution by generating function. They may be solved successively, or by Laplace time-transform. The method of separated solutions is a variant of the latter. It will be used later, when we consider the equations associated with the Random Walk. In any case, we obtain

$$P(n\,|\,\ell,t)=e^{-\gamma t}\,\frac{(\gamma t)^{\ell-n}}{(\ell-n)!}\quad n\leq\ell\qquad\text{for }\ P(n\,|\,\ell;0)=\delta(n,\ell). \tag{7.12b}$$

In another example, we assume that the probability of the birth of a new member is proportional to the pre-existing population (as we have, earlier, in our systematic development when we chose *a(n) = a n*.) Then, we encounter an interesting variant of the pure-birth process, a *Yule* process. Setting $\gamma=0$ in our more general treatment, and absorbing "*a*" into the time-variable, we obtain

$$\frac{\partial G(z\mid \ell;\, t)}{\partial t}+\beta\, z\,(1-z)\frac{\partial G(z\mid \ell;\, t)}{\partial z}=0\,, \tag{7.13}$$

which is the statement that $G(z,t)$ remains constant on the characteristic curves $\frac{dz}{dt}=\beta\, z\,(1-z)$, or $\frac{z}{1-z}e^{-\beta t}$ = constant. Thus, $G(z,t)=\phi(\ \frac{z}{1-z}e^{-\beta t}\ )$, with the arbitrary function ϕ to be determined from initial conditions. Should we begin with a population of *k* elements, so that $P(n\mid \ell,0)=\delta(\ell,k)$, and $\phi(\frac{z}{1-z})=z^k$, we infer $\phi(x)=(\frac{x}{1+x})^k$. Thus,

$$G(z\mid k;\, t)=\left[\frac{z\,e^{-\beta t}}{1-z(1-e^{-\beta t})}\right]^k \tag{7.14}$$

and use of the "negative binomial expansion"

$$(1-x)^{-k}=\sum_{m=0}^{\infty}\binom{m+k-1}{m}x^m$$

brings us to the final

$$P(n\mid k,t)=\binom{n-1}{n-k}e^{-k\beta t}\ \ (1-e^{-\beta t})^{n-k} \tag{7.15}$$

for the probability of finding *n* elements at a time *t* later. Of course, the probability is zero for *n* less than *k*. The equation for the generating function informs us (via differentiation) that the average population (first moment) grows as $\mu_1=e^{\beta t}$, the second moment is $\mu_2=k\,e^{\beta t}[\,(k+1)e^{\beta t}-1]$, and that the dispersion, $\sigma^2=k\,e^{\beta t}[\ e^{\beta t}-1]$, grows rapidly. In this, and other examples, the average and the dispersion may be extracted directly from the equation for the generating function, without solving the equation. One simply differentiates and sets *z* = 1.

Now we shall generalize our system to include deaths (or captures) as well, via a model of *Yule* type, *a(n) =a n*. The equation for the generating function is (recall Eqs. (7.9) and (7.10).

$$\frac{\partial G(z\mid \ell;\, t)}{\partial t}+(\beta\, z-\gamma)(1-z)\frac{\partial G(z\mid \ell;\, t)}{\partial z}=0\,, \tag{7.16),}$$

and computation quite similar to what has gone before leads us to

$$G(z\mid k;\, t)=\left[\frac{\theta(1-e^{-(\beta-\gamma)t})-z(\theta-e^{-(\beta-\gamma)t})}{(1-\theta\,e^{-(\beta-\gamma)t})-z(1-e^{-(\beta-\gamma)t})}\right]^k \tag{7.17}$$

with $\theta = \gamma / \beta$, the ratio of "death" to "birth" parameters. The special case $\beta = \gamma$ gives

$$G(z \mid k;\, t) = \left[\frac{\beta t + z(1-\beta t)}{(1+\beta t) - z\beta t} \right]^k . \tag{7.18}$$

The probability of extinction of the population $P(0 \mid k,t) = G(0 \mid k,t) = \left[\frac{\beta t}{(1+\beta t)}\right]^k$ appears significant, perhaps daunting, until one realizes that in a typical chain reaction the number *k* is very, very large.

The probability for a population of *n* elements—our having begun with *k*—may be extracted from the generating function by binomial expansion. If we introduce

$$r(t) = \frac{1 - e^{-(\beta-\gamma)t}}{1 - \theta e^{-(\beta-\gamma)t}} \quad , \qquad h(t) = \frac{\theta - e^{-(\beta-\gamma)t}}{1 - e^{-(\beta-\gamma)t}}$$

we find

$$P(n \mid k,t) = \sum_{m=0}^{(k,n)} \binom{k}{m} \binom{n+k-m-1}{n-m} \theta^{k-m} [r(t)]^{n+k-m} [-h(t)]^m ,$$

an expression not particularly transparent. (The upper limit signifies the smaller of the quantities (*k*,*n*).) If we begin, at $t = 0$, with a single element, we obtain the simpler

$$P(0 \mid 1,t) = \theta\, r(t) = \theta \frac{1 - e^{-(\beta-\gamma)t}}{1 - \theta e^{-(\beta-\gamma)t}} \quad ,$$

$$P(n \mid 1,t) = (1-\theta)^2\, e^{-(\beta-\gamma)t} \frac{[1 - e^{-(\beta-\gamma)t}]^{n-1}}{[1 - \theta e^{-(\beta-\gamma)t}]^{n+1}} \qquad n > 0. \tag{7.19}$$

Of course, moments can be obtained directly from the equation for the generating function by differentiating its differential equation. We find that the mean and the dispersion travel as

$$\mu_1(t) = k\, e^{(\beta-\gamma)t} \qquad \sigma^2 = \mu_2 - \mu_1^2 = k \left(\frac{\beta+\gamma}{\beta-\gamma}\right) e^{(\beta-\gamma)t} \left(e^{(\beta-\gamma)t} - 1 \right), \tag{7.20}$$

the probability for a particular number of elements increasing, then fading away. The case $\beta = \gamma$ resembles that of a symmetric random walk, but in a half-infinite interval. We find that the mean (position) remains unaltered, and that the half-width increases as the square root of time—as we would expect for a diffusion process. Note, too, that there is a definite probability of extinction for the population. That probability approaches unity as time unfolds if $(\gamma / \beta) > 1$, and approaches (γ / β) if the ratio is less than unity.

Finally, if we allow for any number of offspring per birth, we may write

$$p(n,m)=\gamma\,\delta(n-m+1)+\beta[\,\theta_1\,\delta(n-m-1)+\theta_2\,\delta(n-m-2)+...], \qquad (7.21)$$

with the understanding that γ is the probability of death, β of birth, ($\beta+\gamma=1$) . Then, following the blessed event, θ_j is the probability of j offspring ($\sum_j \theta_j = 1$), and $<j>=\sum_j j\theta_j$, the mean number. The "little generating function" becomes

$$g(z)=\gamma+\beta[\theta_1 z^2+\theta_2 z^3+\theta_3 z^4...] \qquad g(1)=1\ .$$

From it come two quantities regulating the evolution of the population,

$$g^{\otimes}\equiv g'(1)-1=<j>\beta-\gamma \quad \text{and} \quad g''(1)=\beta<j(j+1)>.$$

Differentiation of the equation for the generating function, and setting $z = 1$, leads to

$$\frac{d}{dt}<n>+(1-g'(1))<n>=0$$

$$\frac{d}{dt}<n(n-1)>+2(1-g'(1))<n(n-1)>=g''(1)<n>, \qquad (7.22)$$

an easy pair of equations to solve. Should we begin with k members we will find mean and dispersion

$$<n>=k\,e^{g^{\otimes}t}\ , \qquad \sigma^2\equiv<n^2>-<n>^2=k\left(\frac{g''(1)-g^{\otimes}}{g^{\otimes}}\right)e^{g^{\otimes}t}\left(e^{g^{\otimes}t}-1\right)$$

where $g^{\otimes}=g'(1)-1=<j>\beta-\gamma$ describes the growth rate. In more conventional notation we have a simple and elegant pair of expressions,

$$<n>=k\,e^{(<j>\beta-\gamma)t]} \qquad \sigma^2=k\,\frac{<j^2>\beta+\gamma}{<j>\beta-\gamma}\,e^{(<j>\beta-\gamma)t}\left(e^{(<j>\beta-\gamma)t}-1\right) \qquad (7.23)$$

Note that there is a certain deceptiveness in our results. For example, suppose that $<j>\beta-\gamma=0$, $d<n>/dt = 0$. An observer concerned only with mean values would conclude that one had a stable, uninteresting population. However, the variance is increasing linearly, and the probability of extinction increases with time. The constancy of $<n>$ and the overall normalization of probability is maintained in the face of extinction by a small, "high- n" tail of the distribution. In the special case that birth produces a single offspring, the mathematics of the system closely resembles the one-dimensional random walk. When $\beta=\gamma$ we have a symmetric random walk on a half-infinite interval (see Lecture 8). Since we know that the walker will return to an arbitrary point, we may conclude that extinction will occur in a finite time. In the continuum limit

we encounter a solution to the time-dependent diffusion equation on the half-interval. Then, the appropriate boundary condition at the origin is that there be no return current; the exiting current describes the flow into extinction.

Lecture 8

The Random Walk

The *random walk* may be seen as a special example of a master equation. To quote B. D. Hughes, it may be described ". . . in a . . . frivolous manner as the mathematical theory of the aimless wanderings of an idealized drunkard, . . . yet the applications of the basic formalism to physical problems are numerous and significant and many mathematical results beautiful in their own right may be derived."This topic is rich, both in applications and in intrinsic beauty; the literature associated with it is considerable. The walk may be approached in simple probabilistic terms, as we shall initially, or as a special example of a master equation of type: "discrete times and discrete states."

In its simplest form we consider the walk to be composed of *N* steps along an infinite line (a lattice of points $x_n = na$, actually.) The steps are independent; a step to the right is associated with probability *p*, to the left with probability *q = 1-p*. We seek the probability *P(n,N)* of finding the walker (particle) leaving the origin and arriving at point *n*, after *N* steps, "*r*" to the right and "*l*" to the left. Clearly, this is an example of binomial statistics. In those terms, a step to the right is "success," to the left, "failure." Then, $n = r - l$, $N = r + l$, and the probability we seek is

$$P(n,N) = p^{\frac{1}{2}(N+n)}\, q^{\frac{1}{2}(N-n)} \binom{N}{\frac{1}{2}(N+n)} \qquad -N \le n \le N \,. \tag{8.1}$$

$$= 0 \qquad \text{otherwise}$$

The probability can be retrieved, in a longer fashion, by using a discrete variant of the trick we noted earlier. Thus,

$$P(n,N) = \sum_{\sigma_1} \cdots \sum_{\sigma_N} p(\sigma_1)\, p(\sigma_2) \ldots p(\sigma_N)\; \delta\Big(n - \sum_{m=1}^{N} \sigma_m\Big) \tag{8.2}$$

with

$$\delta(m) = \frac{1}{2\pi} \int_0^{2\pi} d\phi\, e^{i m \phi} \,.$$

Then, quite analogous to our discussion of random flights, we find

$$P(n,N) = \frac{1}{2\pi} \int_0^{2\pi} d\phi\, e^{in\phi} \left[\, p\, e^{-i\phi} + q e^{i\phi} \,\right]^N \tag{8.3}$$

and the binomial expansion of the integrand brings us to Eq. (8.1). What is elegant about this representation is that if we were to consider a symmetric random walk on a lattice of points (n_1, n_2) , we would find

$$P(n_1,n_2,N)=\frac{1}{4^N}\ \frac{1}{2\pi}\int_0^{2\pi}\ d\phi_1\,\frac{1}{2\pi}\int_0^{2\pi}\ d\phi_2\,e^{i(n_1\phi_1+n_2\phi_2)}\,[\ e^{-i\phi_1}+e^{-i\phi_2}+e^{i\phi_1}+e^{i\phi_2}\ ]^N . \tag{8.4}$$

The "4" denotes the number of nearest neighbors, the probability of jumping to each being ¼. The generalization to lattices of higher dimensions is obvious.

Let us return to the symmetric walk along a line and consider the probability distribution when *N* is large. Again, we go from binomial to Gaussian. Consider the distribution after many jumps. We can analyze as before, replacing ℓ by *n*/2 to get

$$P(n,\,N)\sim\sqrt{\frac{2}{\pi N}}\ \exp[-\frac{n^2}{2N}] \tag{8.5}$$

under the assumption that $\frac{n^2}{N}$ is held fixed, and is *O(1)*, as *N* is made large. Thus, the estimate does <u>not</u> hold for points close to the origin.

Note that if we wish to regard *n* as varying continuously, Eq. (8.5) is not normalized correctly. That is because we have "over-counted." Recall that if *N* is even/odd, *n* must be even/odd, else the probability vanishes. In the continuous limit, we must divide by two to get the normalized Gaussian,

$$P(n,\,N)\sim\frac{1}{\sqrt{2\pi N}}\ \exp[-\frac{n^2}{2N}]\ . \tag{8.6}$$

Next, we connect with a simple diffusion process. Let us introduce a spatial scale through $x = n\,a$, a time scale through a "jump frequency" $N=\nu\,t$, and a probability density $\overline{P}(x,t)\,dx=P(n,N)\,dn$, to get

$$\overline{P}(x,t)=\frac{1}{\sqrt{4\pi Dt}}\ \exp[-\frac{x^2}{4Dt}]\ ,$$

where *D* is the diffusion coefficient $D=\frac{a^2\nu}{2}$. This function is the solution to the diffusion (or heat) equation,

$$\frac{\partial\,\overline{P}(x,t))}{\partial t}=D\frac{\partial^{\,2}\overline{P}(x,t)}{\partial x^2}\qquad\overline{P}(x,t=0)=\delta(x)\,. \tag{8.7}$$

Our treatment carries over easily to two and three dimensions.

The Master Equation

A quite different approach introduces both new ideas and a simple approach to the equations which describe asymptotic behavior. Consider the conditional $\pi(n|m)$, which is the probability that a particle (flea?), resident on point *m*, will, on command, jump to point *n*. Now we are describing the random walk as a stochastic process with transition probability $\pi(n|m)$ and master equation

$$P(n,\ N+1) = \pi(n|n+1)\,P(n+1,N) + \pi(n|n-1)P(n-1,N) \tag{8.8}$$

$$= q\,P(n+1,N) + p\,P(n-1,N).$$

Eq. (8.8) is a difference equation which has a solution

$$P(n,N) = p^{\frac{1}{2}(N+n)}\, q^{\frac{1}{2}(N-n)} \binom{N}{\frac{1}{2}(N+n)}. \tag{8.9}$$

Now write $p = \frac{1}{2} + \varepsilon$, $q = \frac{1}{2} - \varepsilon$ and arrange the equation to read

$$P(n,\ N+1) - P(n,\ N) = \frac{1}{2}[\,P(n+1,\ N) - 2\,P(n,\ N) + P(n-1,N)] - \varepsilon[\,P(n+1,\ N) - P(n,\ N)]. \tag{8.10}$$

But this is the discrete, or difference, version of the differential equation,

$$\frac{\partial P(n,N)}{\partial N} + \varepsilon \frac{\partial P(n,N)}{\partial n} = \frac{1}{2}\frac{\partial^2 P(n,N)}{\partial n^2}, \tag{8.11}$$

a "drift-diffusion" equation. With $N = \nu t$, $x = n a$, $\eta = \varepsilon a \nu$, $D = (1/2)\nu a^2$, we have

$$\frac{\partial \bar{P}(x,t)}{\partial t} + \eta \frac{\partial \bar{P}(x,t)}{\partial x} = D\frac{\partial^2 \bar{P}(x,t)}{\partial x^2}, \tag{8.12}$$

an equation for a probability density. One would expect the transition to the continuous picture to be justified in the limit of large *N* and *n*. The earlier approach gave a more precise estimate of the limiting procedure.

The flexibility of the model may be appreciated if we think about polymer chains constructed of segments of length a. One merely replaces "step" with "segment" in the analysis. A one-dimensional image is a bit strained—the chain being folded back and

forth along a line—but, again, the analysis is extended easily to higher dimensions. The polymer version of the equation describing large-*N* behavior is

$$\frac{\partial P(x,N)}{\partial N} = \frac{a^2}{2}\frac{\partial^2 P(x,N)}{\partial x^2}, \tag{8.13}$$

$P(x,N)\,dx$ being the probability that a polymer molecule of *N* links, whose tail is anchored at $x = 0$, will find its head lying between *x* and $x + dx$.

Escape?

When we consider random walk on lattices of dimension greater than one, a question arises naturally. Will a walker, released at some point on a lattice, ever return, and, if never, does it walk out to infinity? The question evokes a rich discussion and we shall give a taste of it. For the simple walks, we begin at the origin.

Recall that *P(n, N)* is the probability of reaching point *n* after *N* jumps. Let us denote the probability of being at the origin after *N* jumps as *P(N)*. These jumps may describe quite complex, looping, paths. Consider the quantity *F(N)*, the probability that the walker, after *N* jumps, returns to the origin for the first time (*first passage)*. We have $P(0) = 1$; and, as a harmless convention, we take $F(0) = 0$.

We will need the generating functions associated with these probabilities, namely

$$\begin{aligned} \bar{F}(z) &= \sum_{N=0}^{\infty} F(N)\, z^N \\ \bar{P}(z) &= \sum_{N=0}^{\infty} P(N)\, z^N . \end{aligned} \tag{8.14}$$

The quantity $\bar{F}(z=1)$ is the probability that the origin will ever be revisited. The defining series is certainly convergent for |z|=1. The convergence of the *P*-series on the circle is not assured. The key point of the analysis is to recognize that a walk returning to the origin after *N* steps will be compounded of a first $F(k)$, *k*-step loop followed by a $P(N-k)$ loop. If we sum over all possibilities, we have

$$P(N) = \sum_{k=1}^{N} P(N-k)\,F(k) \qquad N \geq 1, \tag{8.15}$$

which, upon multiplication by z^N, summing, and using $F(0) = 0$, produces

$$\bar{P}(z) = 1 + \bar{P}(z)\,\bar{F}(z) \qquad \bar{F}(z) = 1 - \frac{1}{\bar{P}(z)}, \tag{8.16}$$

where the limit $z \to 1$ is of particular interest.

Let us begin with the 1-D case: We sum the *P(n, N)* to get, for the generating function,

$$\bar{P}(z) = \frac{1}{2\pi}\int_0^{2\pi} d\phi \frac{1}{1 - z\cos\phi}, \tag{8.17}$$

which diverges as z approaches unity. Thus $\bar{F}(z \to 1) = 1$ and the walker surely returns to
the origin, again and again. In the case of the two-dimensional square lattice, we have

$$\bar{P}(z) = \frac{1}{2\pi}\int_0^{2\pi} d\phi_1 \frac{1}{2\pi}\int_0^{2\pi} d\phi_2 \frac{1}{1 - \frac{z}{2}(\cos\phi_1 + \cos\phi_2)}, \tag{8.18}$$

and for the simple cubic (with six nearest neighbors)

$$\bar{P}(z) = \frac{1}{2\pi}\int_0^{2\pi} d\phi_1 \frac{1}{2\pi}\int_0^{2\pi} d\phi_2 \frac{1}{2\pi}\int_0^{2\pi} d\phi_3 \frac{1}{1 - \frac{z}{3}(\cos\phi_1 + \cos\phi_2 + \cos\phi_3)} . \tag{8.19}$$

These integrals may be collapsed if we represent the denominator by

$$\int_0^{\infty} dt\, e^{-At} = \frac{1}{A}$$

and recognize the Bessel function $I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} d\phi\ e^{x\cos\phi}$. Then,

$$\bar{P}(z) = \int_0^{\infty} dt\, e^{-t}\, I_0^2\left(\frac{tz}{2}\right) \quad \text{(2-D)} \qquad \bar{P}(z) = \int_0^{\infty} dt\, e^{-t}\, I_0^3\left(\frac{tz}{3}\right) \quad \text{(3-D).} \tag{8.20}$$

and we resolve the issue of convergence as z approaches unity by noting that
$I_0(z) \sim \frac{e^z}{\sqrt{2\pi z}}$ for large z with Re(z) > 0. We conclude that the 2-D integral diverges, the 3-D does not. Thus in 3- and greater dimensioned lattices, the walker does not always return home.

A bit of history: In 1939 the integral defining $\bar{P}(1)$ was evaluated, in terms of elliptic integrals, by G. N. Watson in a remarkable tour de force. For the simple, cubic 3-D lattice the probability of return is 0.340, for the body-centered cubic it is 0.282 and for the face-centered, 0.256.

Solving the Difference Equation

There are interesting situations in which the random walk takes place on a finite, or half-infinite interval. *P(n, N)* is then obtained by direct solution of the difference equation. We shall illustrate one approach with the model *p=q=1/2* .

The equation

$$P(n,\ N+1) = \frac{1}{2}[P(n+1,N)+P(n-1,N)] \tag{8.21}$$

will be treated as we often treat partial differential equations, by seeking an appropriate combination of separable solutions. Here, we choose the form *V(N)W(n).* When this trial is entered into Eq. (8.21) we obtain

$$V(N+1)/V(N) \ = \ [W(n+1)+W(n-1)]/2W(n) = \beta^2 ,$$

β^2 being the traditional "separation constant." Then $V(N) = V(0)\beta^{2N}$ and

$$W(n+1) - 2\beta^2 W(n) + W(n-1) = 0 .$$

But the latter equation has exponential solutions, of form λ^n. If—in anticipation of later developments—we write the separation constant as $\beta^2 = \cos\phi$, we find solutions $\lambda_1 = e^{i\phi},\ \lambda_2 = e^{-i\phi}$. These will not diverge in the infinite or semi-infinite lattice. Then, it is natural to consider the superposition

$$P(n,N) = \int_0^{2\pi} \frac{d\phi}{2\pi} \cos^N\phi\,[A(\phi)e^{in\phi} + B(\phi)e^{-in\phi}] . \tag{8.22}$$

In a first example, consider a walk on the infinite lattice beginning at $n = k$, with $N = 0$. Clearly if $A(\phi), B(\phi)$ are chosen so that

$$P(n,N;\ k) = \int_0^{2\pi} \frac{d\phi}{2\pi} \cos^N\phi\ \cos(n-k)\phi\ , \tag{8.23}$$

we have the desired solution. One can check by setting $k = 0$ and using the binomial expansion, to regain Eq. (8.1).

In the second, we consider the half-infinite lattice $n \geq 0$ and imagine the point $n = 0$ to be a sink, whence there is no return. To translate this situation into a boundary condition, we note that the presence of a sink implies

$$P(1,\ N+1) = \frac{1}{2}P(2,N)\ . \tag{8.24}$$

This condition implies, in turn, that $[A(\phi) + B(\phi)]$ vanishes. After setting the initial condition we have

$$P(n, N; k) = \int_0^{2\pi} \frac{d\phi}{\pi} \cos^N \phi \ \sin n\phi \ \sin k\phi \ , \tag{8.25}$$

an expression that is correct for k, $n > 0$. Having the solution, we'd like a measure of the flow into the sink or, from another point of view, the probability that a gambler, beginning with a stake = k, will be ruined after N bets. The answers involve the interesting quantity $P(n=0, N; k)$ which we get from (1/2) $P(1, N-1; k)$. We may speed the evaluation by writing

$$P(n, N; k) = \int_0^{2\pi} \frac{d\phi}{2\pi} \cos^N \phi \ [\cos(n-k)\phi - \cos(n+k)\phi] \ , \tag{8.26}$$

$$= \frac{1}{2^N} \left\{ \binom{N}{\frac{1}{2}(N+n-k)} - \binom{N}{\frac{1}{2}(N+n+k)} \right\}, \tag{8.27}$$

the last equality holding for $-k \le n-k \le N$ in the first expression, $0 \le n+k \le N$ in the second. We have superposed two "infinite lattice" solutions using the familiar method of images. The techniques we have used may be extended, easily, to the case of the asymmetric random walk, and the walk on a finite lattice—in one-dimension, of course.

Lecture 9

Compound Distributions and Branching Processes

It is said that the topic of branching processes arose during an attempt to apply probabilistic notions to a study of the extinction of family names in England. The challenge, then, is to apply our ideas to successive generations of a population. In physics and chemistry we might be following populations in chain-reacting systems, generation by generation, or attempting to understand the particle showers which follow a nuclear disintegration at high energy.

To begin, recall that in these lectures we have frequently been interested in the distribution of the sum of N independent, "identical" variables, X_i, each having distribution $p_1(x)$ and generating function $g(s)$. Suppose that N, too, is governed by a probability distribution, $p(N)$, whose generating function is $H(s)$. We now encounter the issue of *compound distributions.* Suppose further that the outcomes x are non-negative integers, $k = 0, 1, 2, \ldots$ Then, clearly

$$P(k) = \sum_N p(N)\ \Pr\Big(\sum_{i=1}^{N} X_i = k\Big). \tag{9.1}$$

Here, Pr(…) is a conditional probability, $p(k \mid N) \equiv \Pr\big(\sum_{i=1}^{N} X_i = k\big)$ ("Given N, …"), while $p(N)$ is the probability of N.

The generating function for $P(k)$ is

$$G(s) = \sum_n s^k P(k) = \sum_N p(N) \sum_n s^k\, p(k \mid N). \tag{9.2}$$

But we have encountered the function on the rhs before, in Lecture 1. It is

$$\sum_k s^k\ \Pr\Big(\sum_{i=1}^{N} X_i = k\Big) = \sum_k s^k\ p(k \mid N) = g(s)^N. \tag{9.4}$$

This relation is of particular importance in the discussion which follows. From it we infer

$$G(s) = \sum_N p(N)\, g(s)^N = H(g(s)). \tag{9.5}$$

We present two examples. In the first, suppose that we irradiate biological material, creating damaged cells. These cells may, subsequently, be repaired. We ask for the probability of *n* ultimate survivors. Suppose that the radiation produces *N* damaged cells following a Poisson distribution. Thus $H(s) = Exp[\lambda(1-s)]$, with $\lambda =< N >$. Then there is the probability *p* of remaining damaged, *q=1-p* of being repaired, so *g(s) = q+ps*. Overall, then, $G(s) = Exp[\lambda\, p(1-s)]$ and the distribution of survivors (mutations?) is again Poisson, but with a modified expectation.

The second example is that of the famous branching process. Suppose that we begin with a single stochastic object, capable of producing with probability $p(k)$ k= 0, 1, 2 , . . . offspring. The associated generating function is *g(s)*. These offspring independently produce offspring according to the same law. Let us follow this, generation by generation, with a somewhat expanded notation. Thus, $p_1(k_1) \equiv p(k_1)$ is the probability of finding k_1 offspring in the first generation, $p_2(k_2)$ the probability of finding k_2 in the second, and so on. Then, introduce the conditional probability, $p(k_2 \mid k_1) \equiv \Pr(\sum_{i=1}^{k_1} X_i = k_2)$ that the k_1 independent offspring will produce k_2 offspring in the next generation. Thus,

$$p_2(k_2) = \sum_{k_1} p_1(k_1)\; p(k_2 \mid k_1) \tag{9.6}$$

describes the distribution in the second generation. Generally,

$$p_{n+1}(k_{n+1}) = \sum_{k_n} p_n(k_n)\; p(k_{n+1} \mid k_n) \tag{9.7}$$

Now, to the generating functions: If $G_n(s) = \sum_{k_n} s^{k_n} p_n(k_n)$, and we recall that $G_1(s) \equiv g(s)$, we see that

$$\begin{aligned} G_2(s) &= \sum_{k_2} s^{k_2} p_2(k_2) = \sum_{k_2} s^{k_2} \sum_{k_1} p(k_2 \mid k_1)\, p_1(k_1) \\ &= \sum_{k_1} g(s)^{k_1} p_1(k_1) = g(g(s)), \end{aligned} \tag{9.8}$$

where we have used Eq. (9.4). In a similar manner,

$$G_3(s) = \sum_{k_2} g(s)^{k_2} p_2(k_2) = G_2(g(s)) = g(g(g(s))), \tag{9.9}$$

and so on.

Thus we describe successive generations. An easy, immediate result is that if the fundamental distribution has mean value <*k*>, the mean value of the population at the *n*th generation is $< k >_n = < k >^n$. We see this by induction and differentiation, noting that

$$G_{n+1}(s) = g(G_n(s)) \,, \qquad G_n(s=1) = 1 \,. \tag{9.10}$$

For example,

$$< k > = [\frac{dg(s)}{ds}]_{s=1} \,, \quad < k >_2 = [\frac{dG_2(s)}{ds}]_{s=1} = [\frac{dg(g(s))}{ds}]_{s=1} = g'(g(s=1))\, g'(s=1) = < k >^2 \,.$$

Thus, if $< k >$ is less than unity, the population will surely be extinguished in time. If $< k >$ is greater than unity there may be a non-zero probability of extinction but, overall, the population will increase exponentially.

Let us now ask for $p_n(0) = g(g(g(...0)))$, the probability of extinction at the *n*th generation. (We assume, throughout that $0 < p_1(0) < 1$.) We first note that $p_n(0)$ increases with *n*. This is rather obvious, and may be seen by drawing a graph of the function $G_1(s) \equiv g(s)$ for $0 \le s \le 1$. (See Fig. 1.) *g(s)* is a positive, increasing function, and that feature is the key to the analysis. Since $p_1(0) = g(0) > 0$ we will have the progression $p_2(0) = g(g(0)) > g(0) = p_1(0)$, $p_3(0) > p_2(0)$ and so on. Induction completes the proof; $p_n(0)$ continues to increase towards unity. If it approaches unity ultimately, extinction will ultimately occur. This is the case of $< k > = [\frac{dg(s)}{ds}]_{s=1} < 1$. If $p_n(0)$ approaches a smaller limit—and let us denote that value by *x*—that quantity, *x=g(g(g(…x)))*, must be the smallest root of the equation $x = g(x)$. We have encountered a *fixed point.* The situation is displayed in Fig. 1; either <*k*> < 1 and extinction occurs with probability one, or <*k*>>1 and the mean diverges, though there remains a non-zero probability of extinction.

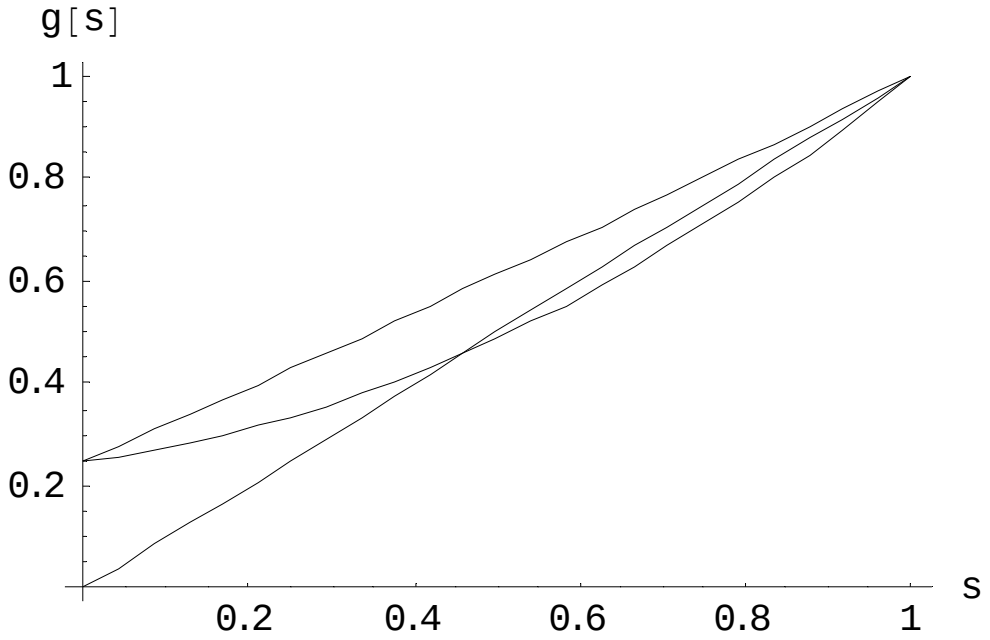


Figure 1. Branching Processes

Two generating functions are compared. They represent two different probability distributions, $p(k)$ for producing offspring. For both, $g(0)$, the probability of extinction in the first generation is 0.25. The curve $g(s) = s$ is included for reference. One distribution (upper curve) leads, ultimately, to extinction. The other (lower curve), by crossing $g(s) = s$, indicates a fixed point in the iteration and a non-zero probability that the population will survive.

Lecture 10

Stochastic Differential Equations

When we discussed Markov stochastic processes, $p_1(x;t)$, the distribution that characterizes the stochastic quantity, *X(t)*, appeared as the solution of an equation of conventional form—an integro-differential or a differential equation. The equation was characterized by a transition probability, suggesting that the process be viewed as a sequence of jumps, these being induced by random interactions with, or in, a host system. Once the transition probability is chosen, the solution of the equation proceeds by conventional means. But there is another way of generating $p_1(x;t)$. It is, perhaps, more direct and appealing than the encounter with Chapman-Kolmogorov. It is an obvious generalization of the Langevin approach, and produces a *stochastic differential equation,* which is another species of animal.

Recall that in considering Brownian motion, we begin with a deterministic equation and replace the term describing the rapidly varying collisional force with a stochastic variable. Here is where we encounter a stochastic differential equation. We anticipate that we will choose a realization of the stochastic process, solve the equation and average later, with respect to some ensemble of functions. The fact that the stochastic term is additive, appearing as a source-, or driving-term in the differential equation, eases the calculations. Often, when we alter deterministic equations with a similar replacement, the stochastic variable appears as a "multiplier." Life is then harder. A classic example is the propagation of waves (e.g., acoustic, electro-magnetic) through random media, where we may encounter

$$[\frac{\partial^2}{\partial x^2} + \frac{\omega^2}{c^2} n^2(x)] E(x,\omega) = 0,$$

displaying a stochastic index of refraction, *n(x)*. Such equations are, obviously, more difficult to analyze than those of Langevin type. We illustrate with an example.

The Randomly Modulated Oscillator:

The deterministic equation is

$$\frac{dx}{dt} - i\omega(1+\varepsilon b(t))x(t) = 0 \qquad x(0) = 1 \tag{10.1}$$

or

$$x(t) = e^{i\omega t}\,[1 + i\varepsilon\omega \int_0^t dt_1\, b(t_1)\, x(t_1)]\,. \tag{10.2}$$

Suppose that the oscillator frequency wobbles symmetrically about ω in a very complicated manner. The wobble is of amplitude ε, which will be small. We choose to treat the behavior statistically. Thus,

$$\frac{dX(t)}{dt} - i\,\omega(1+\varepsilon\,B(t))X(t) = 0 \qquad X(0) = 1\ , \tag{10.3}$$

with the capitalized quantities stochastic, and $<B(t)> = 0$. We have a stochastic process with an associated distribution $p_N(x_1,t_1\,;\,x_2,t_2;\ldots\,x_N,t_N)$. It is driven by our choice of *B(t).* As before, we may regard that quantity as a collection of local and correlated variables, labeled by the time variable, or as an ensemble of functions defined on $0 \le t < \infty$. Usually, we cannot hope to determine more than a few of the moments of the distribution of *X*. It would be pleasant to have self-contained equations for each of the moments, $< X(t)^n >$, but these are, generally, unavailable.

Our almost trivial model, Eq. (10.3), is useful because it may be integrated, and we may then judge the accuracy of the several schemes of approximation we will discuss. Thus,

$$X(t) = e^{i\omega t}\,Exp[\,i\varepsilon\omega\int_0^t dt_1\,B(t_1)]\ .$$

We take $B(t)$ to be Gaussian, whereupon its integral $Z(t) \equiv \int_0^t dt_1\,B(t_1)$ is also Gaussian. Then, the average, $< X(t) >= e^{i\omega t}\ < Exp[\,i\varepsilon\omega\,Z(t)]>$, is simply

$$< X(t) >= e^{i\omega t}\,Exp[-\frac{1}{2}\varepsilon^2\omega^2\,\sigma^2(t)]\ , \tag{10.4}$$

$$\sigma^2(t) = \int_0^t dt_1 \int_0^t dt_2\ < B(t_1)\,B(t_2) > .$$

Often, the correlation may be represented, usefully, as

$$< B(t_1)\ B(t_2) >= C(t_1 - t_2) = \ Exp[-\frac{|\,t_1 - t_2\,|}{T}]. \tag{10.5}$$

Then we find

$$\sigma^2(t) = 2\ T^2\,[\,e^{-\frac{t}{T}} - (1 - \frac{t}{T})]\,.$$

The picture, then, is of an oscillating amplitude damped, on average, by its fluctuation in frequency. In a condensed notation,

$$< X(t) >= e^{i\omega t}\,Exp[-\alpha^2\ [\,e^{-\tau} - (1-\tau)]]\ , \tag{10.6}$$

with $\alpha^2 = \varepsilon^2\omega^2T^2$, $\tau = t/T$.

The damping, whose exponential is proportional to t^2 at early times, and to *t* at late times, is not caused by absorption, which is absent from the model. Rather, its source is "phase mixing," a situation encountered often in the study of plasma waves.

Usually, we are unable to solve the stochastic equation. The options we then face are to approximate the solution to the true equation or to solve, truly, an approximate equation. The first approach involves expanding in some small parameter. If we iterate Eq. (10.2) or (10.3) we obtain

$$X(t)=e^{i\omega t}\,[1+i\omega\varepsilon\int_0^t dt_1\,B(t_1)-\omega^2\varepsilon^2\int_0^t dt_1\int_0^{t_1} dt_2\,B(t_1)B(t_2)+...],$$

$$<X(t)>=e^{i\omega t}\,[1\ -\omega^2\varepsilon^2\int_0^t dt_1\int_0^{t_1} dt_2<B(t_1)B(t_2)>+...]. \qquad (10.7)$$

Thus, naive expansion in the small parameter ε leads to a series in powers of $\varepsilon\omega$ and functions of time that are unbounded. This non-uniform representation is not satisfactory. Retention of an acceptable number of terms gives an expression useful only for relatively short times; we are simply expanding the exponential in Eq. (10.4).

We might try, in a second approach, to construct an equation for <*X(t)*> alone. It will be approximate but, we hope, useful. To begin, direct averaging leads to

$$\frac{d<Y(t)>}{dt}-i\omega\varepsilon<B(t)Y(t)>=0 \qquad <Y(0)>=1\ , \qquad (10.8)$$

where we have introduced $Y(t)=\ e^{-i\omega t}X(t)$ to ease calculation. The product term poses a challenge; we would like to replace it with something containing <*Y(t)*>alone. While simple decoupling, $<B(t)Y(t)>\approx<B(t)><Y(t)>$ leads to nonsense, it does suggest an approach: Let us write

$$Y(t)=<Y(t)>+(Y(t)-<Y(t)>)=PY(t)+QY(t)\,. \qquad (10.9)$$

P and *Q* may be thought of as linear operators; *P* as averaging, and *Q* as measuring the degree of fluctuation (wobble). Clearly, $P^2=P,\ \ Q^2=Q,\ \ PQ=QP=0,\ \ P+Q=1.$ Remembering that *PB(t)* = 0, we may rewrite Eq. (10.8) as

$$\frac{d}{dt}PY(t)-i\omega\varepsilon\,PB(t)\,QY(t)=0 \qquad PY(0)=1 \qquad (10.10)$$

and add its partner

$$(\frac{d}{dt}-i\omega\varepsilon\,QB(t))\ QY(t)=i\omega\varepsilon\,B(t)\,PY(t), \qquad QY(0)=0\ . \qquad (10.11)$$

It appears that we can solve Eq. (10.8) for *QY* in terms of *PY* and substitute into Eq. (10.7) to obtain an equation for <*Y(t)*> alone. While that is indeed so, formally, and no approximation has been made, it is no surprise that the result is a daunting, intractable equation. But a scheme of approximation presents itself−to expand in the *Q's*, in the magnitude of fluctuation. The process is called *smoothing.* In lowest order we neglect the term in Eq. (10.8) that is quadratic in *Q*. We get, then

$$\frac{d}{dt}< Y(t) > +\omega^2\varepsilon^2\ PB(t)\int_0^t dt_1 B(t_1) < Y(t_1) >= 0 \qquad < Y(0) >= 1\ , \qquad (10.12)$$

or, upon averaging,

$$\frac{d}{dt}< Y(t) > +\omega^2\varepsilon^2\ \int_0^t dt_1\ C(t,t_1) < Y(t_1) >= 0 \qquad < Y(0) >= 1\ , \qquad (10.13)$$

a most interesting equation. Since "memory" is present in the integral term, we realize that the stochastic process describing the evolution of *X(t)* cannot be Markovian. (And we have not assumed that the B-statistics are Gaussian.) Should the correlation function be of form $C(t-t_1)$, stationary statistics, the equation will lend itself to easy solution via Laplace transformation. Thus, with

$$\begin{pmatrix}\eta(s)\\ C(s)\end{pmatrix}=\int_0^\infty dt\ e^{-st}\begin{pmatrix}<Y(t)>\\ C(t)\end{pmatrix}, \text{ we get } \ \eta(s)=\frac{1}{s+\varepsilon^2\omega^2 C(s)}\ . \qquad (10.14)$$

The simple, exponential version of the correlations leads to

$$\frac{1}{T}\eta(s)=\frac{1+Ts}{Ts(1+Ts)+\varepsilon^2\omega^2\ T^2} \qquad (10.15)$$

and the damping is exponential,

$$< X(t) >= e^{i\omega t}\, e^{-\frac{\tau}{2}}\,[Cosh(\beta\tau/2)+\frac{Sinh(\beta\tau/2)}{\beta}]\ , \quad \beta=\sqrt{1-4\alpha^2} \qquad (10.15)$$

in the notation used earlier. The smoothing result is quite useful when $1-4\alpha^2>0$. Although the representations (10.15) and (10.6) are different, they assume the same exponential form at late times and, overall, produce numerical values that are surprisingly close. (See Fig. 2). The smoothing approximation gives greater damping. As $T\to 0$, which we may regard as the white noise limit, the damping vanishes. In this example, the lowest order smoothing approximation is quite good. When it is applied to more complicated models its success is not quite as impressive.

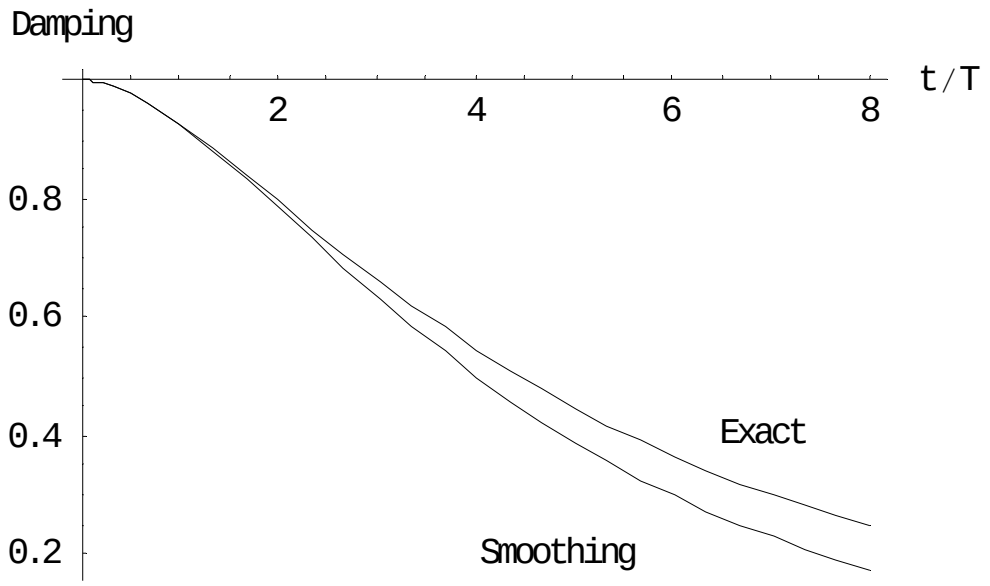


Figure 2. Stochastic Differential Equations

The exact (upper) and the approximate (lower) expressions for the decay of the amplitude of a randomly modulated oscillator, in a "typical" case, described in the text. The approximate expression is that given by "first-order smoothing."

Appendix A

The Maximum Entropy Scheme

The issue is, how do we obtain a priori probabilities to describe the various scenarios, the outcomes of a particular experiment? We have noted the use of symmetry or, more generally, the Principle of Insufficient Reason. A more powerful scheme is based upon the notion of *entropy.* Suppose an experiment has *n* independent outcomes, $i = 1, 2, 3, \ldots n$. The experiment is carried out *N* times, and outcome *i* occurs N_i times. Associated with these outcomes are the frequencies $f_i = N_i / N$ and these, in the limit of large *N*, will become the associated probabilities.

Why entropy, which we define as

$$S = -\sum_{i=1}^{n} f_i \log f_i \ ? \qquad \text{(A1)}$$

Recall, from Lecture 2, that a multiplicity, $W[N_i]$, is associated with a run of *N* giving results N_i. The multiplicity is

$$W[N_i] = \frac{N!}{N_1!\, N_2! \ldots N_n!} \ ; \qquad \text{(A2)}$$

it quantifies the number of scenarios, sequences, which will yield the same set of frequencies. The entropy may be seen to be the Stirling, large-*N*, approximation to the logarithm of the multiplicity. When multiplicity is maximized, so is entropy. And the key point is that when *N* is large, the particular set $\{ N_i \}$ which maximizes the multiplicity, does so overwhelmingly. The vast majority of the possible scenarios have frequencies close to those which maximize the entropy. They are Nature's choice.

In detail, and with some examples: In the simplest situation we would maximize the entropy subject to the obvious constraint $\sum_{i=1}^{n} f_1 = 1$. If we use the technique of Lagrange multipliers, we are led quickly to uniformity, to equal likelihood of every outcome, $f_i = 1/n$. The situation in which we have additional knowledge, expressed in the form of, say, a linear constraint, is more interesting. Suppose that we have $\sum_i \varepsilon_i f_i = E$ in addition. Then, we should seek the extremum of

$$S_1 = \sum_{i=1}^{n} f_i \left[\log f_i + \lambda + \beta \varepsilon_i\right]. \tag{A3}$$

Differentiating with respect to f_k produces

$$\log f_k + (\lambda + 1) + \beta \varepsilon_k = 0 \tag{A4}$$

which, after a little algebra, yields the frequencies/probabilities

$$f_k = \frac{e^{-\beta \varepsilon_k}}{\sum_i e^{-\beta \varepsilon_i}}, \quad E = \frac{\sum_i \varepsilon_i e^{-\beta \varepsilon_i}}{\sum_i e^{-\beta \varepsilon_i}}, \tag{A5}$$

the unknown β being determined by the second relation. Obviously, in this example we have derived the familiar statistical thermodynamics of a closed system of non-interacting "atoms," each of which has *n* states. The entropy principle may be extended to treat problems involving image reconstruction, time series, and communication. It has been expounded most effectively by E. T. Jaynes, whom we have followed here.

Appendix B

The Wiener-Khinchin Relation

Consider a real, fluctuating amplitude, viewed as a stochastic process which is stationary and homogeneous. Then, referring back to Lecture 5,

$$< X^2 >= \int dx\, P(x)\; F(x,0)\; F(x,0) = \int dx\, P(x)\; x^2 \; . \qquad \text{(B1)}$$

Now let us do a Fourier expansion in the equation for the correlation function,

$$< X(t_1)X(t_2) >= \int dx_1 \int dx_2 \;\; p_2(x_1,t_1;x_2,t_2)\; x_1\, x_2 \; = \int dx\, P(x)\; F(x,t_1)\; F(x,t_2) \; .$$

Then,

$$< X(t_1)X(t_2) >= \frac{1}{(2\pi)^2} \int d\omega_1 \int d\omega_2 \exp[i(\omega_1 t_1 + \omega_2 t_2)] \; \{ \int dx\, P(x) \overline{F}(x,\omega_1)\; \overline{F}(x,\omega_2) \} \; .$$

Because the system is homogeneous and stationary, the correlation function depends only upon the difference of the two times. To ensure this, the quantity in curly brackets must be proportional to a delta function, as, say

$$\int dx\, P(x)\; \overline{F}(x,\omega_1)\; \overline{F}(x,\omega_2) \; = 2\pi\, \delta(\omega_1 + \omega_2)\; I(\omega_1) \, . \qquad \text{(B2)}$$

Then, with $t_1 - t_2 = \tau$ we have our main result,

$$< X(\tau)X > = \frac{1}{2\pi} \int d\omega \exp(i\,\omega\tau)\, I(\omega) \, , \qquad < X^2 > = \frac{1}{2\pi} \int d\omega\, I(\omega) \, . \qquad \text{(B3)}$$

The interpretation is: When we perform a harmonic analysis of a fluctuating quantity $X(t)$, $\frac{1}{2\pi} I(\omega)\, d\omega$ is the intensity contained in the frequency interval $d\omega$. Then, $I(\omega)$ and the (auto-) correlation function are Fourier transforms of each other. Should $I(\omega)$ be constant, I_0, we will have $< X(\tau)X > = I_0\, \delta(\tau)$. This is the case of "white noise."

Appendix C

The Fokker-Planck Equation

We begin with

$$T(x_3, x_1; t+\tau) = \int dx_2\, T(x_3, x_2; \tau) T(x_2, x_1; t)\,.$$

We multiply by an arbitrary, smooth function and integrate, to get

$$\int dx_2\, R(x_2) T(x_2, x_1; t+\tau) = \int dx_3 \int dx_2\, R(x_3) T(x_3, x_2; \tau) T(x_2, x_1; t)\,.$$

We expand $R(x_3) = R(x_2) + \sum_{n=1} \frac{1}{n!} \frac{\partial^n R(x_2)}{\partial x_2^n} (x_3 - x_2)^n$ and do the x_3 integration to get

$$\int dx_2\, R(x_2)[T(x_2, x_1; t+\tau) - T(x_2, x_1; t)] = \sum_{n=1} \frac{1}{n!} \int dx_2\, \frac{\partial^n R(x_2)}{\partial x_2^n} T_n(x_2; \tau) T(x_2, x_1; t)\,,$$

with moments, $T_n(x_2; \tau) = \int dx_3\, (x_3 - x_2)^n\, T(x_3, x_2; \tau)$.

After partial integration we have

$$\int dx_2\, R(x_2)[T(x_2, x_1; t+\tau) - T(x_2, x_1; t)] = \sum_{n=1} \frac{(-)^n}{n!} \int dx_2\, R(x_2) \frac{\partial^n}{\partial x_2^n} T_n(x_2; \tau) T(x_2, x_1; t)\,.$$

And now the crucial supposition—that the first two moments are, for small jumps, proportional to the jump-size, τ , and that all others are smaller, of order o(τ) . The latter may be neglected when we pass to the limit, $\tau \to 0$. Then, with

$$T_1(x_2; \tau) = D_1(x_2)\,\tau\,,\ T_2(x_2; \tau) = D_2(x_2)\,\tau,$$

we are led, in the limit, to

$$\int dx_2\, R(x_2)[\frac{\partial}{\partial t} T(x_2, x_1; t) + \frac{\partial}{\partial x_2} D_1(x_2) T(x_2, x_1; t) - \frac{1}{2} \frac{\partial^2}{\partial x_2^2} D_2(x_2) T(x_2, x_1; t)] = 0\,.$$

Since this is true for "all" sufficiently smooth *R(x)*, we may conclude that the quantity in square brackets vanishes. And that is the equation of A. D. Fokker and M. Planck. The quantity D_1 is known as the *drift coefficient*, D_2 the *diffusion coefficient*. The equation states that, under certain circumstances, knowledge of two moments of a function suffices to determine the entire function. It has something of the flavor of a Gaussian approximation.

References

1. Coffey, W. T., Kalmykov, Yu. P., Waldron, J. T.: *The Langevin Equation* (World Scientific, Singapore 1996)

2. Feller, W. : *An Introduction to Probability Theory and its Applications, 2d Edition* (Wiley, New York 1957)

3. Frisch, U. : "Wave Propagation in Random Media" (in *Probabilistic Methods in Applied Mathematics*, ed. by A.T. Bharucha-Reid, Academic Press, New York 1968)

4. Gardiner, C. W. : *Handbook of Stochastic Methods, 2d Edition* (Springer, Berlin 1985)

5. Harris, T. E. : *The Theory of Branching Processes* (Dover, New York 1989)

6. Hughes, B. D. : *Random Walks and Random Environments* (Oxford, New York 1995)

7. Jaynes, E. T. : *On the Rationale of Maximum-Entropy Methods* (Proc. IEEE, vol.70, (#9) pp 939-952, 1982)

8. Khinchin, A. I. : *Mathematical Foundations of Statistical Mechanics*. (Dover, New York 1949)

9. Papoulis, A. : *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York 1965)

10. Risken, H. : *The Fokker-Planck Equation, 2d Edition* (Springer, Berlin 1989)

11. Van Kampen, N. G. : *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam 1981)

12. Wax, N. : *Selected Papers on Noise and Stochastic Processes* (Dover, New York 1954). This collection contains classic papers by Chandrasekhar, Uhlenbeck, Rice, Kac, and Doob.