Probability and random processes

Introduction

Probability has got itself a bit of a bad name in some circles. How could a model that tells you what the answer is likely to be, rather than what it really is, possibly be up to much? Can it be anything better than an admission of defeat? This deterministic prejudice pervades much of physics and engineering, taking Einstein's contention that 'God does not play dice' as its rallying cry. On top of this probability theory does itself very few favours, frequently adopting the guise of either a branch of pure mathematics with axioms, demonstrations and very few applications or a *pot pourri* of combinatorics and conundrums devised to highlight the muddle-headedness of the masses. Fortunately it doesn't have to be like that. In today's session we will review the calculational methods that allow us to characterise the performance of radar systems in terms of detection and false alarm rates and model the noise and clutter processes that degrade the performance. Our approach will be heuristic and motivated by the applications in hand. One or two old favourites, like the prize hidden behind one of several doors and people in a room comparing birthdays, will be mentioned and, with any luck, demystified.

Historically, probability theory grew out of a desire, on the part of the rich and indolent, to gain some advantage in the games of chance with which they filled their leisure hours. Even today this scenario provides us with a useful introduction to the subject. Starting off with simple die and coin tossing games we will develop rules that allow us to characterise their outcomes. We will then generalise these ideas to cover other situations, in particular those where the outcome (e.g. a received signal) takes a continuum of values and develops in time. Combining and processing these outcomes gives us new random variables; how do we characterise their properties? Once again we will move from simple discrete examples to the more complicated situations we encounter in practice. Particular attention will be paid to the distinction between correlated and independent events and the concepts of joint and conditional probabilities. This leads us to Bayes' theorem, which underpins much of the estimation of parameters that characterise observed processes in probabilistic terms. Some commonly used models, including the binomial, Poisson, gamma and normal processes, will be described along the way. We conclude with a brief review of the description of the dynamics of a Gaussian process, in terms of Langevin (or stochastic differential) equations and its power spectrum.

The application of probability theory in practical situations requires us to carry out calculations of varying degrees of complexity; the exercises should provide the practice you might need in doing this. Emphasis will be placed, as it is in the 'real world', on getting useful answers, rather than on the philosophical, psychological and pyrotechnical aspects of the subject.

Heads I win, tails you lose

Perhaps the simplest game of chance is that in which a single coin is tossed and the uppermost face (H or T) is recorded. In any one game one cannot predict the outcome, obviously. If we were to repeat the game many times we would be surprised if there was a great preponderance of either heads or tails. Experience, intuition or whatever would lead us to expect approximately equal numbers of heads and tails and that ultimately, in a very large number of trials, the numbers of heads and tails would be effectively indistinguishable. We 'feel' that heads and tails are equally likely to occur in any one trial and that this quantifiable indeterminism in one event is nonetheless predictably manifest in the outcome of a large number of such events. In the coin tossing game we recognise two outcomes, H and T. Either outcome is one of these two options; we identify the frequency of occurrence of either outcome with the ratio of number of ways it can be achieved to the total number of outcomes of a single trial. This intuitive approach, identifying the outcome of an exercise in enumeration with a frequency or 'probability' provides us with a blueprint from which we can build up an extensive calculational framework. There are many ways of introducing the basic ideas ('axioms') and interpretation of probabilistic calculus, whose conflicting claims of rigour and appeal to common sense do much to make the subject seem magical. We will just carry on regardless, extrapolating from this simple example until we make contact with the problems in which we are interested.

Throwing a single die provides another example: the die has 6 faces corresponding to six outcomes. The probability of any one of these turning up in a single throw is taken to be1/6.

These examples highlight two salient features of a probability: that it is positive (or zero) and that the sum of the probabilities of all possible outcomes of a trial is one.

The problems of enumeration become more complex as the game of chance becomes more involved. Thus, if we tossed two coins, what would be the frequency (probability) of our getting two heads?. The total number of outcomes is 4 (TT,HH,TH and HT) of which just one provides two heads. The probability of such an outcome 'is' 1/4. The probability of throwing a head and a tail (irrespective of order) is 1/2 etc. Throw two dice; what is the probability of scoring six? There are 36 outcomes of which 5 ((1,5), (5,1), (2,4), (4,2), (3,3)) give us the required score. This provides the answer 5/36. You might like to carry through the argument for all the scores between 2 and 12 and check that the sum of the probabilities of all possible scores adds up to 1. More complicated games (the National Lottery, poker, blackjack) can be analysed in much the same way, though considerably more skill is needed in enumerating the possible outcomes efficiently. We don't have to worry too much about the detail of the combinatorial problems that arise. One problem is important though; how many ways can you pick *n* objects from a set of *N*, irrespective of their order and without putting them back?

It could be you (** the Lotto you)**

Pick your first object; there are *N* ways of doing this. Pick your second; there are *N*-1 ways of doing this. Continuing in this way we find that there are

$$N(N-1)(N-2)\cdots(N-n+1) = N!/(N-n)! \qquad (1)$$

ways of picking *n* objects without replacement. This calculation counts sets of objects that differ only in the order in which they are drawn out; if we do not wish to distinguish between these orderings we must remove this multiple counting by dividing by the number of ways in which *n* objects can be ordered i.e. *n*!. For example, the total number of outcomes of the National Lottery draw is

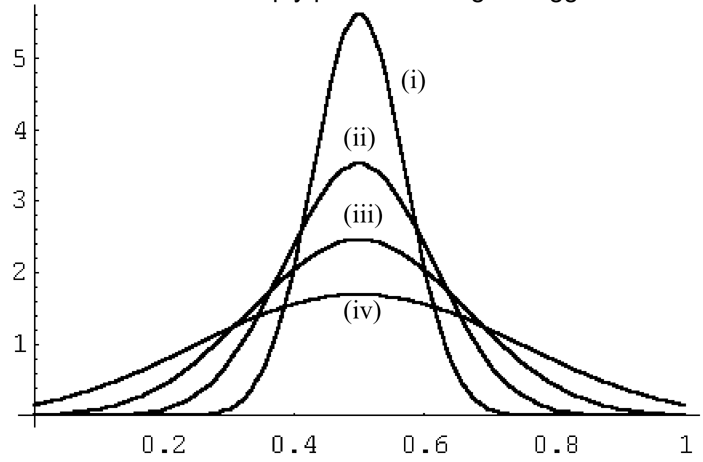
$$\frac{49!}{6!43!} = 13{,}983{,}816 \qquad (2)$$

Your single pound buys one of these, so that the probability of your winning the Jackpot is fairly small. Try calculating the probabilities of some of the lesser winning combinations and marvel at the mark-up implicit in the prizes they actually dole out for them.

This calculation has a more significant application. If, rather than considering the outcome of a single coin tossing game, we could consider how many heads might be thrown in a sequence of *N* games. So far we have convinced ourselves that it will be about $N/2$. Can we do better than this and calculate the probability of there being *n* heads? Again we seek to enumerate the possible outcomes. All told there are 2^N outcomes to the sequence of *N* games. Of these $N!/(N-n)!\,n!$ will consist of *n* heads. Thus the probability of there being *n* heads thrown in *N* games is given by

$$P(n|N) = \frac{N!}{2^N (N-n)!\,n!}\,. \qquad (3)$$

If we plot this out for several values of *N* we see that it does indeed peak around $n = N/2$, and becomes more sharply peaked as *N* gets bigger.



This plot shows $NP(n|N)$ as a function of n/N , *N*=5,10,20,50

Let's do something a bit different. So far we have just worked out probabilities; now we are going to use these to work out average (mean, expectation) values. If we did a whole lot of *N* fold coin tossings what would be the average value of *n* we would obtain? We work this out as

$$\langle n \rangle_N = \sum_{n=0}^{N} nP(n|N) = \frac{N!}{2^N} \sum_{n=0}^{n} \frac{n}{n!(N-n)!} = \frac{N}{2} \tag{4}$$

This is reassuring. Now let's work out the mean square value of *n*

$$\left\langle n^2 \right\rangle_N = \sum_{n=0}^{N} n^2 P(n|N) = \frac{N!}{2^N} \sum_{n=0}^{n} \frac{n^2}{n!(N-n)!} = \frac{N(N+1)}{4} \tag{5}$$

Other expectation values can be worked out similarly. The ones we have here give us a good idea of the fluctuations of *n* about its mean value; a commonly used measure is the normalised variance

$$\frac{\left\langle n^2 \right\rangle_N - \langle n \rangle_N^2}{\langle n \rangle_N^2} = \frac{1}{N} \tag{6}$$

which we see gets smaller as the number of games per trial gets larger, all in accord with the intuitive picture we outlined at the beginning.

So far, so good. However there is more to life than tossing unbiased coins. At the very least we could cheat. More commendably we could seek to extend this analysis to other binary outcome experiments where the outcomes are not the same (e.g. is there a target there or not) Rather than construct a set of outcomes that incorporate the bias in the tossed coins we seek to combine probabilities of individual events more efficiently. If two events E_1, E_2 are mutually exclusive outcomes of a trial then the probability that one or the other will occur is given by

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \tag{7}$$

If two independent events E_1, E_2 have probabilities $P(E_1), P(E_2)$ then the probability of their both occurring is given by

$$P(E_1, E_2) = P(E_1)P(E_2) \tag{8}$$

(These can be regarded as definitions, axioms or sort of obvious, depending on your taste). If we toss a biased coin the probability of a head popping up is *p* (As head and tail exhaust all possible outcomes of the trial the probability of a tail is 1-*p*). So, if we toss *N* coins and obtain a configuration of *n* heads the probability of that given configuration arising is $p^n(1-p)^{N-n}$. However there are $N!/(N-n)!\,n!$ such configurations, each with *n* heads. So to get the probability of any of these, equally probable and mutually exclusive, combinations occurring in the trial we form

$$P(n|N) = \frac{p^n(1-p)^{N-n} N!}{(N-n)!\,n!} \tag{9}$$

It is reassuring to note that our unbiased coin result is a special case of this and that

$$\sum_{n=0}^{N} P(n|N) = 1 \tag{10}$$

as we would hope.

A couple of probabilistic 'party tricks'

The technique of appropriate enumeration of outcomes allows one to demonstrate a variety of rather 'counter-intuitive' probabilistic results whose appeal (in my estimation at least) is based more in psychology than in mathematics. A classic problem of this type asks 'how many people do you have to have in a room before there is a 50% chance that (at least) two of them have the same birthday?' Subconscious solipsism suggests that you think of this in terms of how many would you have to have in the room before one of them had the same birthday as yourself; there could be millions in there before that happened. A moment's disinterested consideration shows that if, forgetting leap years, there are 366 people in a room then at least two of them must have the same birthday. Viewed in this light the problem seems more sensible. Actually working out the answer requires low cunning. The probability that at least two people have the same birthday is equal to 1 minus the probability that everyone in the room has a different birthday. 365 days in a year, *N* chaps in the room: total number of outcomes= 365^N Total number of ways of *N* people having all different birthdays: 365!/(365-*N*)!. Probability of *N* people all having different birthdays: 365!/(365-*N*)! 365^{-N}. Probability of at least two people in the room having the same birthday: 1-365!/(365-*N*)! 365^{-N}.Typical values of this number are

{2, 0.00273973}, {4, 0.0163559}, {6, 0.0404625}, {8, 0.0743353},

{10, 0.116948}, {12, 0.167025}, {14, 0.223103}, {16, 0.283604},

{18, 0.346911}, {20, 0.411438}, {22, 0.475695}, {24, 0.538344},

{26, 0.598241}, {28, 0.654461}, {30, 0.706316}, {32, 0.753348},

{34, 0.795317}, {36, 0.832182}, {38, 0.864068}, {40, 0.891232},

{42, 0.91403}, {44, 0.932885}, {46, 0.948253}, {48, 0.960598},

{50, 0.970374}

So twenty three people in a room give you a 50% chance of at least two of them having the same birthday.

An altogether more subtle (and notorious) problem invites you to consider a TV game show in which a contestant is placed in front of three doors and told, quite truthfully, that the prize of his dreams lies behind one of them. He is invited to pick a door behind which he feels the prize awaits. The smiling host then opens one of the other doors to (and) reveal(s) nothing at all. Now, he asks the contestant, what are you going to do? You can change your mind if you want, or you can stick with your original choice. What should this poor fellow do? To make his best move the contestant has to consider two scenarios: (i) the host knows where the prize was and has opened a door that he knew had nothing behind it, and, (ii) the host opened one of the two doors (not chosen by the contestant) at random and unwittingly revealed that there was nothing behind it. Systematic enumeration of cases comes to our aid. Label the doors 1,2,3. Without loss of generality we can assign the label 1 to the door behind which the prize resides; disappointment lurks behind doors 2 and 3. If the host knows where the prize is things pan out as follows:

> (i) Our man has chosen door 1, the host opens another door (under these circumstances there is no distinction between 2 and 3);
> (ii) Our man has chosen door 2, the host opens door 3;
> (iii) Our man has chosen door 3, the host opens door 2.

Consequently, by changing his choice to the other unopened door (not initially chosen by him) the contestant has a 2/3 chance of getting the prize; if he stays put he has 1/3 chance. If, however, the host's revelation was the result of a random choice between the two remaining doors, the contestant must consider the following options:

> (i) he chose 1, the host opened 2;
> (ii) he chose 1, the host opened 3:
> (iii) he chose 2, the host opened 3,
> (iv) he chose 3, the host opened 2:

these outcomes are equally probable and are consistent with the observed events. Under these circumstances he has an equal chance of getting the prize if he stays put (2 out of 4: (i) and (ii)) or if he changes his choice (2 out of 4: (iii) and (iv)). As he does not know what the host is up to he has, at the very least, nothing to lose by changing his choice to the other unopened door. Much of the distress caused by this party piece has its origin in the sufferer's unwitting inability to distinguish between the scenario in which the host knows where the prize lies, and acts accordingly, and that in which he does not. For what it's worth, Paul Erdos thought that this problem was wholly counterintuitive and beyond his capacity to understand at a visceral level; he could, of course do all the sums involved.

Limiting forms of the binomial distribution

So far our coin tossing game has introduced us to the 'frequentist' view of probability, evaluation of these probabilities by direct enumeration of cases and by combining the probabilities of constituent events and the calculation of expectation values. The binomial distribution has furnished us with an explicit expression for the probability of the number of occurrences of a given outcome to a succession of N independent binary trials. At this point it is interesting to consider what happens to this distribution in the limit of a very large N. To investigate such limiting behaviour we make use of Stirling's approximation to the factorials in the binomial coefficient. Two quite different limiting forms, with different physical interpretations, can be derived. In the first of these we retain discrete values of n and allow p to scale with N. Thus we have

$$
\begin{aligned}
P(n|N) &= \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}; \quad p = \frac{r}{N}; \quad N \to \infty \\
P(n|N) &\sim \frac{N^n}{n!}\left(\frac{r}{N}\right)^n \left(1-\frac{r}{N}\right)^{N-n} \sim \exp(-r)\frac{r^n}{n!}
\end{aligned} \tag{11}
$$

This is known as the Poisson distribution and is characterised by the single parameter *r*. Another, rather illuminating, approach to the Poisson distribution is to consider the probability $p_n(t)$ that, in a time *t*, *n* counts of a randomly occurring process (e.g. arrival of cosmic rays, shot noise etc.) . The equations of motion for *p* and the associated initial conditions embodying a conservation of probability, are

$$
\begin{aligned}
&\frac{dp_n(t)}{dt} = \alpha\left(p_{n-1}(t) - p_n(t)\right) \\
&\frac{dp_0(t)}{dt} = -\alpha p_0(t) \\
&p_n(0) = \delta_{n,0}
\end{aligned} \tag{12}
$$

It is easy to solve these differential equations (using the Laplace transform is probably quickest) to obtain

$$
p_n(t) = \frac{(\alpha t)^n}{n!}\exp(-\alpha t)\,. \tag{13}
$$

This identifies the Poisson distribution with that of a population subject to random immigrations, the parameter α being identified with their mean rate of arrival.

Another limiting form of the binomial distribution emerges when both *n* and *N* become large, so that the former can be thought of as a continuous variable. In our derivation of the Poisson distribution we used Stirling's approximation in its crudest form; in the present case we will have to be a little bit more careful. We set $n = xN$ and invoke the approximation

$$
\log z! \sim (z + 1/2)\log(z) + \log(2\pi)/2 - z \tag{14}
$$

to give

$$
\begin{aligned}
\log\left(P(n|N)\right) \sim\ & Nx\log p + N(1-x)\log(1-p) + (N+1/2)\log N - \left(N(1-x)+1/2\right)\log\left(N(1-x)\right) \\
& - (Nx + 1/2)\log(xN) - 1/2\log(2\pi)
\end{aligned} \tag{15}
$$

A quick look at the earlier figure suggests that we seek some sort of Gaussian approximation to *P*, best achieved by expanding its logarithm about its stationary value. So, by setting

$$
\frac{\partial \log P(n|N)}{\partial x} = 0\,, \tag{16}
$$

we find that this maximum is located at

$$
x_m = p + O(1/N)\,; \tag{17}
$$

to generate the quadratic term in the expansion we note that

$$\left.\frac{\partial^2 \log P(n|N)}{\partial x^2}\right|_{x=x_m} = -\frac{N}{p(1-p)}\,. \tag{18}$$

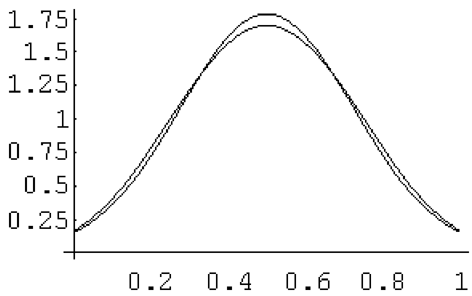
This leads us to

$$\log\left(P(n|N)\right) \approx -1/2\log(2\pi) - 1/2\log\left(Np(1-p)\right) - \frac{(n-Np)^2}{2Np(1-p)} \tag{19}$$

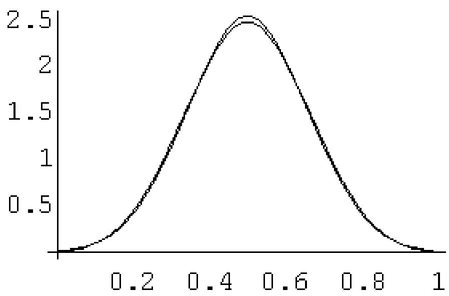
or

$$P(n|N) \approx \frac{1}{\sqrt{2\pi Np(1-p)}}\exp\left(-\frac{(n-Np)^2}{2Np(1-p)}\right) \tag{20}$$

which has the expected Gaussian form. A comparison with plots of the binomial distribution shows that this approximation is not bad for *N* =5 and jolly good for *N* =10



N=5



N=10

The connection between this continuous limit of the binomial distribution and random walks and the diffusion process suggests itself, and is made more explicit in exercise 2.

Continuously varying random quantities

Our coin-tossers' introduction has focused our attention on outcomes with discrete values and provided an illustration of the basic properties that we might expect of a probability (positivity, normalisability, use in calculation of expectation values). It is relatively straightforward to extend these formal concepts to an outcome that can take a continuum of values *x*. We do this by defining a probability distribution *F* and its associated probability density *P*. (These two things are rather different but have similar names, the misuse of which is a source of delight to pedants and other sorry fellows.) Thus we define

$F(x) =$ Probability that the random variable takes a value less than or equal to x

Note that $F(-\infty) = 0;\quad F(\infty) = 1$. The probability density is defined by

$$P(x) = \frac{dF(x)}{dx} \tag{21}$$

which has the rather loose interpretation that the probability that the random variable takes values between x and $x + dx$ is given by $P(x)dx$. Sometimes, to avoid confusion *F* and *P* are labelled with both the random variable's symbol (usually **bold**) and the value that it takes.

Thus we might have $P_{\mathbf{x}}(x)$. The probability that the random variable takes a value exceeding some threshold is given by

$$1-F(x_T)=\int_{x_T}^{\infty}P(x)dx \tag{22}$$

expectation values are given in terms of integrals

$$\langle f\rangle=\int_{-\infty}^{\infty}f(x)P(x)dx\ . \tag{23}$$

Two particularly useful classes of expectation values are the moments

$$\left\langle x^n\right\rangle=\int_{-\infty}^{\infty}x^nP(x)dx \tag{24}$$

and the characteristic function

$$C(k)=\left\langle \exp(ikx)\right\rangle=\int_{-\infty}^{\infty}\exp(ikx)P(x)dx \tag{25}$$

which are related through

$$C(k)=\sum_{n=0}^{\infty}\frac{(ik)^n}{n!}\left\langle x^n\right\rangle\ . \tag{26}$$

Just because F has all the properties of a probability, it does not mean that all, or indeed any, of the moments of the corresponding density exist; when this happens you cannot expand the characteristic function in a Taylor series as we have done here. However, for the clutter models we are interested in, all the moments exist and the characteristic function is well behaved. In some cases, where the random variable takes only positive values, it is convenient to define the characteristic function as the Laplace transform of the pdf:

$$C(s)=\left\langle \exp(-sx)\right\rangle\ . \tag{27}$$

Examples of density functions of values taken by a random variable have cropped up along the way throughout the course and so should not surprise you too much; we are also in good shape to knock them into submission analytically. So we recall:

The Gaussian distribution

$$P(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-(x-m)^2/2\sigma^2\right)\quad -\infty<x<\infty\,; \tag{28}$$

The gamma distribution

$$P(x)=\frac{b^\nu}{\Gamma(\nu)}x^{\nu-1}\exp(-bx)\quad 0\le x<\infty\ ; \tag{29}$$

The Cauchy distribution

$$P(x) = \frac{a}{\pi}\frac{1}{a^2 + x^2} \qquad -\infty < x < \infty\,; \tag{30}$$

The beta distribution

$$P(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \qquad 0 \le x \le 1. \tag{31}$$

You may like to work out the moments and characteristic functions of these distributions; armed as you are with material from the earlier sessions, you should have no trouble.

If we have two random variables **x,y** taking values *x,y* then, much as in the single variable case, we can define a joint probability distribution $F_{\mathbf{xy}}(x,y)$, the probability that **x,y** take values less than or equal to *x,y.* (Note that $F(\infty,\infty) = 1$) Using this joint probability distribution we can evaluate the corresponding probability density function

$$P_{\mathbf{xy}}(x,y) \; = \frac{\partial^2}{\partial x \partial y} F(x,y) \tag{32}$$

from which we can determine expectation values

$$\langle f \rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy f(x,y) P_{\mathbf{xy}}(x,y). \tag{33}$$

Given $F_{\mathbf{xy}}(x,y)$ we can find the corresponding probability distribution of values taken by **x** as $F_{\mathbf{xy}}(x,\infty)$; the corresponding probability density function is found by integrating over all possible values of *y* to give the marginal density:

$$P_{\mathbf{x}}(x) = \int_{-\infty}^{\infty} dy P_{\mathbf{xy}}(x,y)\,. \tag{34}$$

In general $P_{\mathbf{xy}}(x,y)$ is not merely the product of the marginal densities of *x* and *y* i.e.

$$P_{\mathbf{xy}}(x,y) \neq P_{\mathbf{x}}(x) P_{\mathbf{y}}(y)\,; \tag{35}$$

if the equality does hold then the variables **x,y** are said to be independent

The correlation function of **x** and **y** is given by

$$\langle xy \rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, xy P_{\mathbf{xy}}(x,y)\,. \tag{36}$$

For correlated variables $\langle xy \rangle \neq \langle x \rangle \langle y \rangle$; when the variables are independent the correlation function factorises: $\langle xy \rangle = \langle x \rangle \langle y \rangle$.

When **x**,**y** are independent we would expect the value taken by one to be unaffected by the value taken by the other; strongly correlated random variables would be expected to show a significant coupling between the values they take. This idea can be formalised in terms of a conditional probability density. The probability of two events occurring (e.g. the outcome HH of a game in which two coins are tossed) is $P(E_1,E_2)$. The probability that the event E_1 occurs, given that the event E_2 has occurred is referred to as the conditional probability is given by

$$P(E_1|E_2) = \frac{P(E_1,E_2)}{P(E_2)}. \tag{37}$$

In the un-biased two coin game we have

$$\begin{aligned} &P(H) = P(T) = 1/2; \quad P(H,H) = P(H,T) = P(T,H) = P(T,T) = 1/4 \\ &P(H|T) = \frac{P(H,T)}{P(T)} = 1/2 \end{aligned} \tag{38}$$

which makes sense; perhaps you would like to construct other simple examples along these lines. This idea can be generalised to the case where the 'events' are continuous random variables. Thus the probability that **x** takes a value less than or equal to *x*, given that **y** takes a value less than or equal to *y* is given by the conditional distribution

$$F_{\mathbf{xy}}(x|y) = \frac{F_{\mathbf{xy}}(x,y)}{F_{\mathbf{y}}(y)}; \tag{39}$$

the corresponding conditional density, giving the probability that **x** taxes a value between x and $x+dx$, given that **y** takes a value between y and $y+dy$ is given by $P_{\mathbf{xy}}(x|y)dx$ where $P_{\mathbf{xy}}(x|y) = \frac{P_{\mathbf{xy}}(x,y)}{P_{\mathbf{y}}(y)}$. Two random variables **x**,**y** are said to be independent if

$$F_{\mathbf{xy}}(x,y) = F_{\mathbf{y}}(y)F_{\mathbf{x}}(x); \quad F_{\mathbf{xy}}(x|y) = F_{\mathbf{x}}(x) \tag{40}$$

or, equivalently,

$$P_{\mathbf{xy}}(x,y) = P_{\mathbf{y}}(y)P_{\mathbf{x}}(x); \quad P_{\mathbf{xy}}(x|y) = P_{\mathbf{x}}(x) \tag{41}$$

As $P_{\mathbf{xy}}(x,y) = P_{\mathbf{xy}}(x|y)P_{\mathbf{y}}(y) = P_{\mathbf{xy}}(y|x)P_{\mathbf{x}}(x)$ the conditional density specifying the 'probability of *x* given *y*' can be related to the 'probability of *y* given *x*' through Bayes' theorem which states that

$$P_{\mathbf{xy}}(x|y) = \frac{P_{\mathbf{xy}}(y|x)P_{\mathbf{x}}(x)}{P_{\mathbf{y}}(y)} = \frac{P_{\mathbf{xy}}(y|x)P_{\mathbf{x}}(x)}{\int P_{\mathbf{xy}}(y|x)dx}. \tag{42}$$

In some circles this result is dismissed as mere algebra, in others it is identified as the fundamental point of contact between probability theory and the real world. At the very least it underpins the systematic detection and estimation of signals in noise

These ideas are readily extended to three and more random variables; perhaps you would like to check this out (Papoulis' book, 'Probability, Random Variables and Stochastic Processes', McGraw Hill, 1965, Section 8.1 is a good reference for this.)

Functions of random variables

If you have a function that maps values of **x** into those of some other random variable **y**, what is the pdf of its value *y*? Determining this is just a matter of changing variables, as long as the function maps one to one. Merely by changing variables we have

$$P_{\mathbf{y}}(y) = P_{\mathbf{x}}(x(y))\frac{dx}{dy}, \tag{43}$$

where $x = x(y)$ is the inverse of the function *y*. For example, consider the pdf of the logarithm of a gamma distributed variable:

$$\begin{aligned} &y = \log(x) \quad x = \exp(y) \quad \frac{dx}{dy} = \exp(y) \\ &P_{\mathbf{x}}(x) = \frac{b^{\nu}}{\Gamma(\nu)} x^{\nu-1} \exp(-bx) \quad 0 \le x < \infty \\ &P_{\mathbf{y}}(y) = \frac{b^{\nu}}{\Gamma(\nu)} \exp(\nu y - b\exp(y)) \quad -\infty < y < \infty \end{aligned} \tag{44}$$

Perhaps you would like to consider what you might do if the mapping is not one to one.

Consider two random variables **x**,**y** from which you form two new random variables **u,v**, with values given by

$$u = u(x,y), \quad v = v(x,y); \tag{45}$$

we assume that this mapping can be inverted through

$$x = x(u,v), \quad y = y(u,v). \tag{46}$$

The joint density of *u*,*v* is given by

$$\begin{aligned} P_{\mathbf{uv}}(u,v) &= P_{\mathbf{xy}}(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \\ &= \frac{P_{\mathbf{xy}}(x(u,v), y(u,v))}{\left| \frac{\partial(u,v)}{\partial(x,y)} \right|} \end{aligned} \tag{47}$$

where the Jacobian of the transformation is given by

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \frac{1}{\frac{\partial(x,y)}{\partial(u,v)}}. \tag{48}$$

If **x**,**y** are combined to form a single new random variable **z** we can find the density its values as follows. We consider the transformation to the new variables **z**, **y**, whose values have the joint density

$$P_{\mathbf{zy}}(z,y)=\frac{P_{\mathbf{xy}}(x(z,y),y)}{\left|\frac{\partial(z,y)}{\partial(x,y)}\right|}=\frac{P_{\mathbf{xy}}(x(z,y),y)}{\left|\frac{\partial z}{\partial x}\right|_{x=x(z,y)}}. \tag{49}$$

The density of the values of **z** is then obtained by integrating over *y*:

$$P_{\mathbf{z}}(z)=\int_Y P_{\mathbf{zy}}(z,y)dy=\int_Y \frac{P_{\mathbf{xy}}(x(z,y),y)}{\left|\frac{\partial z}{\partial x}\right|_{x=x(z,y)}}dy\,. \tag{50}$$

You have to be a bit careful, in carrying out this integration, to ensure that *Y*, the range of integration, is consistent with the value taken by *z*. We will consider an example of this shortly.

We will now look at a couple of particularly important examples. Say we have two independent random variables **x** and **y** that we add together to form **z**. What is the pdf of *z*?. We see from the general formula we have just derived that it is the convolution of the pdfs of the constituent parts:

$$P_{\mathbf{z}}(z)=\int_{-\infty}^{\infty} P_{\mathbf{x}}(z-y)P_{\mathbf{y}}(y)dy \quad \mathbf{z}=\mathbf{x}+\mathbf{y} \tag{51}$$

From the properties of the Fourier and Laplace transforms we see that the characteristic function of *z* is given by the product of the characteristic functions of the *x* and *y*. This result is very useful. Say, for example, you add together two gamma distributed random variables; what is the pdf of the resultant? Now we know that

$$y=x+x'\ (x,x'>0) \qquad P_{\mathbf{y}}(y)=\int_0^y P_{\mathbf{x}}(y-x')P_{\mathbf{x'}}(x')dx' \tag{52}$$

It is an interesting exercise to evaluate the convolution of the pdfs directly, and see how things sort themselves out - if you are keen you can do the same thing with Laplace transforms. Thus, if we add two independent random variables from identical gamma distributions, the pdf of their sum is given by:

$$\begin{aligned}P_y(y)&=\frac{b^{2\nu}}{\Gamma(\nu)^2}\exp(-by)\int_0^y (y-x)^{\nu-1}x^{\nu-1}dx\\&=b^{2\nu}y^{2\nu-1}\exp(-by)\frac{\int_0^1(1-s)^{\nu-1}s^{\nu-1}ds}{\Gamma(\nu)^2}\\&=\frac{b^{2\nu}y^{2\nu-1}\exp(-by)}{\Gamma(2\nu)}\end{aligned} \tag{53}$$

In the second step we have recognised that the integral is expressible in terms of gamma functions; when we do this everything sorts out very nicely. If the random variables are drawn from gamma distributions with identical scale parameters *b*, but different shape parameters ν_1,ν_2 , essentially the same argument goes through to give us

$$P_y(y) = \frac{b^{\nu_1+\nu_2} y^{\nu_1+\nu_2-1} \exp(-by)}{\Gamma(\nu_1+\nu_2)} \tag{54}$$

The most general case, in which both the scale and shape parameters of the two distributions differ, gives us:

$$\begin{aligned} P_y(y) &= \frac{b_1^{\nu_1} b_2^{\nu_2}}{\Gamma(\nu_1)\Gamma(\nu_2)} \int_0^y (y-x)^{\nu_1-1} \exp(-b_1(y-x)) x^{\nu_2-1} \exp(-b_2 x)\, dx \\ &= \frac{b_1^{\nu_1} b_2^{\nu_2} \exp(-b_1 y)}{\Gamma(\nu_1)\Gamma(\nu_2)} y^{\nu_1+\nu_2-1} \int_0^1 (1-s)^{\nu_1-1} s^{\nu_2-1} \exp(-(b_2-b_1)ys) ds \end{aligned} \tag{55}$$

The exponential in the integrand can be expanded and integrated term by term; the integrals arising from the individual terms can all be expressed in terms of gamma functions:

$$\begin{aligned} P_y(y) &= \frac{b_1^{\nu_1} b_2^{\nu_2} \exp(-b_1 y)}{\Gamma(\nu_1)\Gamma(\nu_2)} y^{\nu_1+\nu_2-1} \sum_{t=0}^{\infty} \frac{\left[-(b_2-b_1)y\right]^t}{t!} \int_0^1 (1-s)^{\nu_1-1} s^{\nu_2+t-1} ds \\ &= \frac{b_1^{\nu_1} b_2^{\nu_2} \exp(-b_1 y)}{\Gamma(\nu_1)\Gamma(\nu_2)} y^{\nu_1+\nu_2-1} \sum_{t=0}^{\infty} \frac{\left[-(b_2-b_1)y\right]^t}{t!} \frac{\Gamma(\nu_1)\Gamma(\iota_2+t)}{\Gamma(\iota_2+\nu_1+t)} \\ &= \frac{b_1^{\nu_1} b_2^{\nu_2} \exp(-b_1 y)}{\Gamma(\nu_1+\nu_2)} y^{\nu_1+\nu_2-1} \; {}_1F_1(\nu_2;\nu_1+\nu_2;(b_1-b_2)y) \end{aligned} \tag{56}$$

Here we have identified the confluent hypergeometric function from its series form. The invariance of this expression under the interchange of the shape and scale parameters can be established using Kummer's transformation of the confluent hypergeometric function. In the case where the scale parameters differ but the shape parameters are the same this result reduces to:

$$P_y(y) = \frac{(b_1 b_2)^{\nu}\sqrt{\pi}}{\Gamma(\nu)} \left((b_1-b_2)y\right)^{\frac{1}{2}-\nu} \exp\left(-\frac{(b_1+b_2)y}{2}\right) I_{\nu-\frac{1}{2}}\left(\frac{(b_1-b_2)y}{2}\right). \tag{57}$$

Here I is a modified Bessel function of the first kind; the symmetry of this expression in the two different scale parameters is obvious.

The key point here is that the sum of two gamma distributed random variables, drawn from the same distribution, is itself gamma distributed; the same is true if the distributions have the same scale parameter b but different shape parameters. The preservation of the gamma form of the pdf goes adrift if the distributions of the summed random variables have different scale parameters. One of the exercises has you look at the summation of random variables drawn from Gaussian and Cauchy distributions.

You can do much the same thing for a product of random variables. So if $\mathbf{z} = \mathbf{xy}$ then

$$P_{\mathbf{z}}(z) = \int P_{\mathbf{x}}(z/y) \frac{1}{|y|} P_{\mathbf{y}}(y) dy \tag{58}$$

which is the Mellin convolution of the pdfs of the constituent random variables. What do you think the analogue of the multiplicative property of the characteristic functions is in this case?

The univariate and bivariate normal process

As the normal or Gaussian process pops up all over the place as a statistical model we will spend a little time reviewing its properties. To start with we will consider a single random variable (univariate process) **x** whose pdf is

$$P_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x-m)^2/2\sigma^2\right) \tag{59}$$

It is easy to show that m is the mean value taken by **x**; σ^2 is similarly identified with the variance of the process

$$\begin{aligned} \langle x\rangle &= \int_{-\infty}^{\infty} xP_{\mathbf{x}}(x)dx = m \\ \left\langle (x-\langle x\rangle)^2\right\rangle &= \langle x^2\rangle - \langle x\rangle^2 = \int_{-\infty}^{\infty} x^2 P_{\mathbf{x}}(x)dx - m^2 = \sigma^2 \end{aligned} \tag{60}$$

The characteristic function of the distribution is given by

$$\begin{aligned} \langle \exp(ikx)\rangle &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} dx \exp(ikx)\exp\left(-(x-m)^2/2\sigma^2\right) \\ &= \exp(ikm)\exp\left(-k^2\sigma^2/2\right) \end{aligned} \tag{61}$$

This can be used to show that the sum of two Gaussian variables is itself a Gaussian random variable whose mean is the sum of the means of the two added processes and whose variance is the sum of their variances. You may like to check this directly from the convolution of their pdfs. It follows from this that the sum of any number of Gaussian random variables is itself Gaussian. The moments of the Gaussian distribution can be read off directly from the characteristic function. To tidy things up a little we will let the mean value m be zero. We then find that

$$\langle \exp(ikx)\rangle = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!}\langle x^n\rangle = \sum_{p=0}^{\infty} \frac{\left(-k^2/2\right)^p}{p!}\langle x^2\rangle^p \tag{62}$$

so that, on equating coefficients of powers of k,

$$\langle x^{2n}\rangle = \frac{(2n)!}{n!2^n}\langle x^2\rangle^n . \tag{63}$$

You might like to prove this by direct integration and subsequent use of the Gamma function duplication formula. This result provides an example of the factorisation property of expectation values of products of $2n$ zero mean Gaussian variables: this is equal to the product of the expectation values of pairs of the variables, summed over all possible distinct decompositions of the original product into pairs. (Check the combinatorial aspects of this.) Thus we might have

$$\begin{aligned} &\langle x_1x_2x_3x_4\rangle = \langle x_1x_2\rangle\langle x_3x_4\rangle + \langle x_1x_3\rangle\langle x_2x_4\rangle + \langle x_1x_4\rangle\langle x_2x_3\rangle \\ &\langle x^4\rangle = 3\langle x^2\rangle^2 \end{aligned} \tag{64}$$

The bivariate normal process consists of two random variables **x,y** (for example the **I** and Q components of a signal) whose joint probability density function is

$$P_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma^2\sqrt{(1-r^2)}}\exp\left(-\frac{(x^2+y^2-2rxy)}{2\sigma^2(1-r^2)}\right)\quad -\infty<x,y<\infty . \tag{65}$$

We can find the marginal distributions of *x* and *y* by simple integration:

$$\begin{aligned}P_{\mathbf{x}}(x)&=\int_{-\infty}^{\infty}P_{\mathbf{xy}}(x,y)dy=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-x^2/2\sigma^2\right)\\ P_{\mathbf{y}}(y)&=\int_{-\infty}^{\infty}P_{\mathbf{xy}}(x,y)dx=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-y^2/2\sigma^2\right)\end{aligned} \tag{66}$$

x and **y** are not independent; we can see this straight away from

$$P_{\mathbf{xy}}(x,y)\neq P_{\mathbf{x}}(x)P_{\mathbf{y}}(y). \tag{67}$$

The mean square values of *x* and *y* are

$$\left\langle x^2\right\rangle=\int_{-\infty}^{\infty}x^2P_{\mathbf{x}}(x)dx=\sigma^2=\left\langle y^2\right\rangle \tag{69}$$

Direct evaluation of the correlation function, by an integration that you might like to check, gives us

$$\left\langle xy\right\rangle=\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dyxyP_{\mathbf{xy}}(x,y)=r\sigma^2 . \tag{70}$$

These calculations allow us to interpret the parameters occurring in the original pdf. In many cases **x**, **y** are uncorrelated (e.g. I and Q components measured at the same time) and the pdf reduces to

$$P_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)\quad -\infty<x,y<\infty \tag{71}$$

for which $P_{\mathbf{xy}}(x,y)=P_{\mathbf{x}}(x)P_{\mathbf{y}}(y)$ holds. Another set of random variables that we might be interested in (amplitude and phase) have values defined by

$$\begin{aligned}z&=\sqrt{x^2+y^2}\\ \phi&=\tan^{-1}(y/x)\end{aligned} \tag{72}$$

The associated pdf is

$$P_{\mathbf{z}\phi}(z,\phi)=\frac{z}{2\pi\sigma^2\sqrt{1-r^2}}\exp\left(-\frac{z^2}{2\sigma^2(1-r^2)}(1-2r\cos\phi\sin\phi)\right)\quad 0\le z<\infty,\quad 0\le\phi<2\pi \tag{73}$$

from which we can derive the marginal distributions

$$P_{\mathbf{z}}(z)=\frac{z}{\sigma^2\sqrt{1-r^2}}\exp\left(-\frac{z^2}{2\sigma^2\left(1-r^2\right)}\right)I_0\left(\frac{rz^2}{2\sigma^2\left(1-r^2\right)}\right)$$
$$P_{\mathbf{f}}(\phi)=\frac{\sqrt{1-r^2}}{2\pi}\frac{1}{1-r\sin 2\phi} \qquad (74)$$

You might like to check what these reduce to in the uncorrelated case. As our last example we consider the 'intensity' $I=x^2+y^2$, in the uncorrelated case. What is the pdf of this fellow? We could obviously read this off from the 'envelope' pdf with a quick change in variables. Rather than do this, though, we will illustrate the way in which one has to be a little careful in 'marginalising' the joint pdf (in this case of *I* and *y*) obtained from a change in variables. In our uncorrelated case we have

$$P_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{\left(x^2+y^2\right)}{2\sigma^2}\right) \quad -\infty<x,y<\infty \qquad (75)$$

The joint density in the new variables is

$$P_{\mathbf{Iy}}(I,y)=\frac{1}{\pi\sigma^2}\frac{\exp\left(-I/2\sigma^2\right)}{2\sqrt{I-y^2}} \quad 0\le I<\infty, \quad -\sqrt{I}\le y\le\sqrt{I} \qquad (76)$$

Note that a factor of two has crept in here (for positive and negative values of *x* contributing the same value of x^2), and that the range of *y* is restricted by the condition $y^2\le I$. To determine the marginal distribution of *I* we integrate over *y*:

$$P_{\mathbf{I}}(I)=\int_{-\sqrt{I}}^{\sqrt{I}}P_{\mathbf{Iy}}(I,y)dy=\frac{\exp\left(-I/2\sigma^2\right)}{2\sigma^2}. \qquad (77)$$

The time evolution of random processes.

So far we have said very little about the way a random process develops in time. We will now consider two complementary ways in which this might be done. The first, which is perhaps more familiar to those with a background in physics, is to supplement the equation of motion of the variable by a random term. It is usual practice to describe this process as white noise, which de-correlates instantaneously on the time scale of variations in the random variable. Gaussian factorisation properties are also assumed. Perhaps the simplest Langevin equation would be that describing a freely diffusing particle, whose position is *x*. Thus we might write

$$\frac{dx}{dt}=f(t); \quad \langle f(t_1)f(t_2)\rangle=2A\delta(t_1-t_2) \ . \qquad (78)$$

This we can solve formally as

$$\Delta x(t)=x(t)-x(0)=\int_0^t f(t_1)dt_1 \qquad (79)$$

The mean square value of this displacement is calculated as follows, exploiting the delta function property of the white noise

$$\left\langle \Delta x(t)^2 \right\rangle = \int_0^t dt_1 \int_0^t dt_2 \left\langle f(t_1) f(t_2) \right\rangle = 2A \int_0^t dt_1 \int_0^t dt_2 \delta(t_1 - t_2) = 2At \,. \tag{80}$$

One can show in much the same way that

$$\left\langle \Delta x(t_1) \Delta x(t_2) \right\rangle = 2A \min(t_1, t_2) \tag{81}$$

A Gaussian process evolving in time with a zero mean and the correlation property (81) is called a Wiener process, and can be used to construct a rigorous formulation of the rather cavalier integration over white noise that we are indulging in here. This framework is frequently referred to as the Ito calculus, and is much in vogue in financial modeling and other circles. The assumed Gaussian properties of the white noise are manifest in those of the diffusive displacement; for example its mean fourth power is

$$\begin{aligned} \left\langle \Delta x(t)^4 \right\rangle &= \int_0^t dt_1 \int_0^t dt_2 \int_0^t dt_3 \int_0^t dt_4 \left\langle f(t_1) f(t_2) f(t_3) f(t_4) \right\rangle \\ &= 3(2A)^2 \int_0^t dt_1 \int_0^t dt_2 \int_0^t dt_3 \int_0^t dt_4 \left(\delta(t_1 - t_2) \delta(t_3 - t_4) \right) \\ &= 3 \left\langle \Delta x(t)^2 \right\rangle^2 \end{aligned} \tag{82}$$

As the mean square displacement of a diffusing particle grows with time as $2Dt$ we can identify A in (80) with the diffusion constant D. Now let's consider an over-damped Brownian harmonic oscillator, described by the Langevin equation

$$\frac{dx}{dt} = -\frac{m\omega_0^2}{\zeta} x + f(t). \tag{83}$$

This can be solved using the integrating factor method to give us

$$x(t) = \exp\left(-m\omega_0^2 t/\zeta\right) x(0) + \int_0^t \exp\left(-m\omega_0^2 (t - t_1)/\zeta\right) f(t_1) dt_1 \tag{84}$$

This process is often referred to as the Ornstein Uhlenbeck process. The mean square displacement can be calculated much as before

$$\left\langle \Delta x(t)^2 \right\rangle = x(0)^2 \left(1 - \exp\left(-m\omega_0^2 t/\zeta\right)\right)^2 + \frac{D\zeta}{m\omega_0^2} \left(1 - \exp\left(-2m\omega_0^2 t/\zeta\right)\right) \tag{85}$$

At short times we see that this is dominated by the diffusive behaviour (80), with the effects of the restoring force coming in to quadratic order. At long times we have

$$\left\langle x(t)^2 \right\rangle = \frac{D\zeta}{m\omega_0^2} = \frac{kT}{m\omega_0^2} \tag{86}$$

which we have equated with the Boltzman average of the harmonically bound particle's displacement form the origin. This leads us to the celebrated Einstein relation between friction and diffusion constants, and provides us with an example of a fluctuation dissipation theorem.

$$D = \frac{kT}{\zeta} = \frac{kT}{6\pi\eta a} \,. \tag{87}$$

The correlation function of the particle's position can also be calculated as

$$\begin{aligned}\langle x(t_1)x(t_2)\rangle &= \exp(-\alpha(t_1+t_2))\left(x(0)^2+2D\int_0^{\min(t_1t_2)}\exp(2\alpha t)dt\right)\\ &= \exp(-\alpha(t_1+t_2))\left(x(0)^2-D/\alpha\right)+\frac{D}{\alpha}\exp(-\alpha|t_1-t_2|);\quad \alpha=\frac{m\omega_0^2}{\zeta}\end{aligned} \tag{88}$$

Neglecting transient terms we can write this as

$$\langle x(t)x(0)\rangle = \frac{kT}{m\omega_0^2}\exp\left(-m\omega_0^2|t|/\zeta\right) \tag{89}$$

The conditional probability $P(x,t\,|\,x_0)$ that a diffusing particle, initially at x_0, is found at *x* at time *t*, satisfies the diffusion equation:

$$\frac{\partial P(x,t\,|\,x_0)}{\partial t} = D\frac{\partial^2 P(x,t\,|\,x_0)}{\partial x^2} \tag{90}$$

The corresponding quantity for the OU process satisfies

$$\frac{\partial P(x,t\,|\,x_0)}{\partial t} = \frac{\partial}{\partial x}\left(\frac{m\omega_0^2}{\zeta}xP(x,t\,|\,x_0)\right)+D\frac{\partial^2 P(x,t\,|\,x_0)}{\partial x^2} \tag{91}$$

the so-called Fokker Planck equation. $P(x,t\,|\,x_0)$ can be constructed quite straightforwardly for a Gaussian process. Referring to (65) we see that

$$P_{\mathbf{y|x}}(y\,|\,x) = \frac{P_{\mathbf{x,y}}(x,y)}{P_{\mathbf{x}}(x)}$$

so that

$$P(x,t\,|\,x_0) = \frac{\exp\left(-(x-x_0\rho(t))^2/2\langle x^2\rangle\left(1-\rho(t)^2\right)\right)}{\sqrt{2\pi\langle x^2\rangle\left(1-\rho(t)^2\right)}} \tag{92}$$

When we set

$$\langle x^2\rangle = \frac{\zeta D}{m\omega_0^2};\quad \rho(t)=\exp\left(-\frac{m\omega_0^2 t}{\zeta}\right) \tag{93}$$

we see that $P(x,t\,|\,x_0)$ satisfies (91) along with the initial condition

$$\lim_{t\to 0}P(x,t\,|\,x_0) = \delta(x-x_0). \tag{94}$$

It is possible to extend the description of time varying random processes to include non-linear drifts and state dependent noise powers (volatilities, as the finance boys call them) and to vector processes; an equivalent description in terms of a Fokker Planck equation can also be constructed. Thus we are able to describe a wide variety of non-Gaussian processes. We will revisit stochastic differential equations when we look at simulation and Kalman filters. Should you want to find out a bit more about the Ito calculus and related matters, you might try the first two or three chapters of 'The Mathematics of Financial Derivatives' by Wilmott, Howison and DeWynne, CUP, 1995, or 'Stochastic Differential Equations', Oksendal, Springer Universitext, 2000. (The former also lets you sound off knowledgeably about all that Rogue Trader stuff.)

Another way in which one might discern the temporal properties of a random process is to consider it as a time series, observed over a period of duration T. This record can be used to construct a correlation function as

$$\langle x(t)x(t+\tau)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x(t)x(t+\tau)dt \quad . \tag{95}$$

The Fourier transform of the random variable, again observed over the interval T, is

$$\tilde{x}_T(\omega) = \int_{-T/2}^{T/2} \exp(i\omega t)x(t)dt \tag{96}$$

(We recall that this is straightforward to do in practice, using an FFT) This can be used to construct the correlation function (95) as

$$\begin{aligned}
&\lim_{T\to\infty}\frac{1}{(2\pi)^2}\frac{1}{T}\int_{-\infty}^{\infty}dt\int_{-\infty}^{\infty}d\omega_1\int_{-\infty}^{\infty}d\omega_2\,\tilde{x}_T(\omega_1)\tilde{x}_T(\omega_2)\exp(-it(\omega_1+\omega_2)-i\omega_1\tau)\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega_1\exp(-i\omega_1\tau)\lim_{T\to\infty}\frac{|\tilde{x}_T(\omega_1)|^2}{T}\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega_1\exp(-i\omega_1\tau)S(\omega_1);\qquad S(\omega_1)=\lim_{T\to\infty}\frac{|\tilde{x}_T(\omega_1)|^2}{T}
\end{aligned} \tag{97}$$

Here we have identified the correlation function as the Fourier transform of the correlation of the power spectrum of the process. This fundamental result is known as the Wiener Khintchine theorem. We will see in a couple of sessions' time how this allows us to synthesise a Gaussian process with a specified correlation function, a tactic that is very useful when we come to simulate random processes. As long as we don't bother about rigour we can establish contact between the spectral and Langevin descriptions of the noise process. Fourier transforming the Langevin equation (83) gives us

$$\begin{aligned}
&\tilde{x}(\omega)\left(\frac{m\omega_0^2}{\zeta}-i\omega\right)=\tilde{f}(\omega)\\
&\frac{|\tilde{x}(\omega)|^2}{T}=\frac{\zeta^2}{m^2\omega_0^4+\zeta^2\omega^2}\frac{|\tilde{f}(\omega)|^2}{T}
\end{aligned} \tag{98}$$

The power spectrum for white noise is $2D$ at all frequencies, so we can write the correlation function using the Wiener Khintchine theorem as

$$\langle x(t)x(0)\rangle = \frac{D}{\pi}\int_{-\infty}^{\infty}\frac{\zeta^2\cos(\omega t)}{\zeta^2\omega^2+m^2\omega_0^4}d\omega \tag{99}$$

which reduces to (89) when we take account of the Einstein relation.

The description of a process in terms of a conditional probability satisfying a Fokker Planck equation assumes that it has no memory, i.e. is unaffected by its previous history. Viewed another way, where one goes next depends only on where one is now. Written formally in terms of a conditional probability we have

$$P(x_n,t_n \mid x_{n-1},t_{n-1};x_{n-2},t_{n-2};\cdots x_1,t_1) = P(x_n,t_n \mid x_{n-1},t_{n-1}); \tag{100}$$

we also have

$$P(x,t \mid x_0,t_0) = \int dx' P(x,t \mid x',t')P(x',t' \mid x_0,t_0); \quad t > t' > t_0. \tag{101}$$

These properties are characteristic of a Markoff process, and are frequently assumed in practice. This idea can be extended to a vector process, any given component of which may itself not be Markoff. Modeling memory effects by having extra 'hidden' processes rattling round in the background is a widely used strategy in many applications. A simple example is the Brownian harmonic oscillator discussed in the last of the exercises. The position and velocity of the particle together constitute a vector process that is Markoff; either considered in isolation would not be Markoff. The requirement of Markoff behaviour imposes restrictions on the dynamics of a stochastic variable: a one dimensional Gaussian process that is Markoff necessarily has an exponentially decaying correlation function.

These ideas are easily extended to the complex Gaussian process, where we consider a complex signal

$$\begin{aligned} Z(t) &= E_{\mathrm{I}}(t) + iE_{\mathrm{Q}}(t); \quad -T/2 < t < T/2 \\ &= 0; \text{ otherwise} \end{aligned} \tag{102}$$

The Fourier transform of this signal is given by

$$\hat{Z}(\omega) = \int_{-\infty}^{\infty} dt \exp(i\omega t) Z(t), \tag{103}$$

from which we form the power spectrum as

$$S(\omega) = \frac{\left|\hat{Z}(\omega)\right|^2}{2\pi T} \tag{104}$$

whose interpretation as the power in a given frequency component is evident. The power spectrum can be written in terms of the I and Q components of the signal as follows

$$\begin{aligned} S(\omega) &= \frac{1}{2\pi T} \int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2 \exp(i\omega(t_1 - t_2))(E_{\mathrm{I}}(t_1) + iE_{\mathrm{Q}}(t_1))(E_{\mathrm{I}}(t_2) - iE_{\mathrm{Q}}(t_2)) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t)\left[\langle E_{\mathrm{I}}(0)E_{\mathrm{I}}(t)\rangle + \langle E_{\mathrm{Q}}(0)E_{\mathrm{Q}}(t)\rangle + i\left(\langle E_{\mathrm{I}}(0)E_{\mathrm{Q}}(t)\rangle - \langle E_{\mathrm{Q}}(0)E_{\mathrm{I}}(t)\rangle\right)\right] \end{aligned} \tag{105}$$

The statistical equivalence and independence of the I and Q components of the signal result in

$$\langle E_I(t)E_I(t+\tau)\rangle = \langle E_Q(t)E_Q(t+\tau)\rangle = \psi\rho(\tau)$$
$$\langle E_I(t)E_Q(t+\tau)\rangle = -\langle E_Q(t)E_I(t+\tau)\rangle = \psi\lambda(\tau)$$
$$\langle {E_I}^2\rangle = \langle {E_Q}^2\rangle = \psi = \langle I\rangle/2 \tag{106}$$

$$\rho(\tau) = \frac{\int d\omega \cos(\omega\tau)S(\omega)}{\int d\omega S(\omega)}; \quad \lambda(\tau) = \frac{\int d\omega \sin(\omega\tau)S(\omega)}{\int d\omega S(\omega)}$$

while the integral of the power spectrum over all frequencies yields the total mean power or intensity in the signal

$$\int_{-\infty}^{\infty} d\omega S(\omega) = \langle {E_Q}^2\rangle + \langle {E_I}^2\rangle = \langle E^2\rangle = \langle I\rangle. \tag{107}$$

(The power spectrum and the Doppler spectrum are the same thing) The values of the I and Q components at two different times constitute a four dimensional Gaussian process that can be characterised by their joint pdf, which is in turn specified by the power spectrum. Thus

$$P(E_I(t), E_Q(t), E_I(t+\tau), E_Q(t+\tau)) = \frac{\exp\left(-\frac{1}{2}\tilde{\mathbf{x}}.\mathbf{K}^{-1}.\mathbf{x}\right)}{(2\pi)^2\sqrt{\det\mathbf{K}}} \tag{108}$$

where **x** is the column vector

$$\mathbf{x} = \begin{pmatrix} E_I(t) \\ E_Q(t) \\ E_I(t+\tau) \\ E_Q(t+\tau) \end{pmatrix} \tag{109}$$

$\tilde{\mathbf{x}}$ is its transpose and **K** is the covariance matrix

$$\mathbf{K} = \psi \begin{pmatrix} 1 & 0 & \rho & \lambda \\ 0 & 1 & -\lambda & \rho \\ \rho & -\lambda & 1 & 0 \\ \lambda & \rho & 0 & 1 \end{pmatrix}. \tag{110}$$

It is easy to show that

$$\det\mathbf{K} = \psi^4(1-\lambda^2-\rho^2)^2 = \psi^4(1-k^2)^2 \tag{111}$$

and that

$$\mathbf{K}^{-1} = \left[\psi(1-k^2)\right]^{-1} \begin{pmatrix} 1 & 0 & -\rho & -\lambda \\ 0 & 1 & \lambda & -\rho \\ -\rho & \lambda & 1 & 0 \\ -\lambda & -\rho & 0 & 1 \end{pmatrix}. \tag{112}$$

The keen reader will show that we recover the familiar 2 dimensional Gaussian when we integrate over the **I** and Q components of the field at one time. We can do everything in polar (amplitude and phase) rather than Cartesian (**I** and Q) components. Thus we have the Rayleigh distribution

$$P(E,\theta) = \frac{E}{2\pi\psi}\exp\left(-\frac{E^2}{2\psi}\right); 0 \le E < \infty, 0 \le \theta < 2\pi \tag{113}$$

and the rather more complicated joint distribution of the amplitudes and phases at two different times:

$$\begin{aligned} P(E(t),E(t+\tau),\theta(t),\theta(t+\tau)) &= \frac{E(t)E(t+\tau)}{(2\pi\psi)^2(1-k^2)} \times \\ &\times \exp\left\{--\frac{1}{2\psi(1-k^2)}\left[E(t)^2 + E(t+\tau)^2 - 2\rho E(t)E(t+\tau)\cos(\theta(t+\tau)-\theta(t)-\phi_0)\right]\right\} \end{aligned} \tag{114}$$

where $\phi_0 = \tan^{-1}\left(\frac{\lambda}{\rho}\right)$.

Exercises

The examples this week are intended to illustrate some basic concepts and calculational procedures that occur in the application of probability theory to physical problems. We start off with a typical enumeration of cases exercise that is interesting, though perhaps no longer that topical. The second extends the discussion of the binomial distribution to the treatment of a random walk and its connection with the diffusion process; Stirling's approximation and the diffusion equation crop up here as well. Example 3 considers properties of simple combinations of random variables; we focus on the familiar Gaussian and rather more pathological Cauchy distributions here. Example 4 requires some use of the 'flash' material from the previous session in the solution of a practical problem. Finally you are invited to look a little more closely at the analysis of Brownian harmonic motion, following in the footsteps of the pioneers of the subject. (You might want to look at the classic papers brought together by Wax in '*Selected Papers on Noise and Stochastic Processes*', Dover, New York, 1954, in particular those by Chandrasekhar, Rice, Ornstein and Uhlenbeck and Wang and Uhlenbeck.)

1 Work out the probabilities of getting 0,1,2,3,4,5 and 5+the bonus ball in the Lottery. Show that the sum of the probabilities of getting 0,1,2,3,4,5 and 6 balls is unity, as you would hope. Can you relate this finding to a more general property of the hypergeometric function? Before the punters realised what a rip-off it all was, though possibly without the benefit of the analysis that you are undertaking, the Sunday Times gave details of the number of winners etc each week. Thus, on 30th May 1998 4 people shared a Jackpot of £8M, 21 got £122,640 for a 5+ bonus, 829 got £1,941 for 5 balls, and 52,974 lucky fellows got £66 for 4 balls. Use this information to estimate how much money Camelot was trousering at that time.

2 Consider a 1 dimensional random walk made up of successive steps, of length Δ and temporal duration τ, occurring with equal probability in either a positive or negative direction, starting at time zero at the origin. Show that the probability that, after s steps (i.e. at time $t = s\tau$), the walk is at a point m steplengths from the origin (i.e at $x = m\Delta$) is

$$P(m,s) = \frac{s!}{2^s((s+m)/2)!((s-m)/2)!} .$$

Arguing as in the lecture notes show that, as s,m become large together, we can write

$$P(m,s) \approx \left(\frac{2}{\pi s}\right)^{1/2} \exp(-m^2/2s) .$$

Also show that

$$P(m,s+1) = \frac{1}{2}\left(P(m-1,s) + P(m+1,s)\right)$$

By expressing this recurrence relation in terms of the variables t,x show that in the limit of small Δ, τ it is equivalent to the diffusion equation

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2}; \quad D = \frac{\Delta^2}{2\tau}$$

Can you relate the solution of this partial differential equation, satisfying the initial condition $P(x,0) = \delta(x)$, to the limiting form of $P(m,s)$ derived earlier? (there is a sneaky factor of 2 lurking around in this one) How might this analysis be modified by the presence of a reflecting or absorbing wall?

3 Evaluate the characteristic functions of the Gaussian and Cauchy distributions. Can you expand the latter in a Taylor series and relate this to the moments of the corresponding distribution as we did for the former in the notes?. Using your expressions for the characteristic functions verify that the sums of 2 and n independent Gaussian random variables are each themselves a Gaussian r.v.; what can you say about sums of 2 and n independent Cauchy random variables? Form and directly evaluate the convolutions of 2 Gaussian pdfs and two Cauchy pdfs and compare the results with those obtained by the Fourier inversion of products of the characteristic functions. What are the pdfs of the random variables formed of a product and a quotient of (i) two independent zero mean Gaussian random variables and (ii) two independent zero mean Cauchy distributed random variables? Comment on your results.

4 The joint density of a correlated bivariate normal process is

$$P_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma^2\sqrt{(1-r^2)}}\exp\left(-\frac{(x^2+y^2-2rxy)}{2\sigma^2(1-r^2)}\right)\qquad -\infty<x,y<\infty$$

Derive an expression for $\langle |xy| \rangle$ .Evaluate the pdf of the product xy and investigate its limits as r tends to zero and to unity. (This is quite a neat little sum that, rather surprisingly, does not appear in any of the standard textbooks; the answer is expressed in terms of a K Bessel function.) Derive an expression for $\left\langle |xy|^n \right\rangle$.

5 During the session we considered the over-damped Brownian harmonic oscillator described by the Langevin equation (83). This provides an analysis appropriate to timescales over which fluctuations in position are observed. Velocity fluctuations de-correlate on a significantly shorter time scale. To include these in a Langevin model we first identify the rate of change of x as a velocity y, whose rate of change is in turn affected by a harmonic restoring force, a viscous drag and a white noise driving term.

$$\frac{dx}{dt}=y;\quad \frac{dy}{dt}=-\omega^2x-\zeta y+\eta(t)$$
$$\langle \eta(t_1)\eta(t_2)\rangle=2\kappa\delta(t_1-t_2)$$

Using the rather slapdash method described in the session, find the power spectra of the processes x and y. Write down expressions for the corresponding correlation functions. Identify the constant κ by equating the mean square velocity with its Boltzmann average; the mean square value is obtained by setting t to zero in the appropriate correlation function and confronting a slightly nasty looking integral, which yields to low cunning. Check that this value of κ now leads to the expected mean square value of x. If you can evaluate the correlation functions at non-zero time, so much the better. (A knowledge of contour integration or a table of Fourier transforms like the Bateman Manuscript collection might come in handy here.) Have a look at solving the Langevin equations in the time domain; unless you are really keen just outline what you would do, rather than flogging it all out. Details of this latter approach are given in Chandrasekhar's paper. The slapdash method adopted here, which focuses more attention on the power spectrum and neglects transient effects, is known as the method of Rice, and is applied to Brownian motion in the paper by Wang and Uhlenbeck.