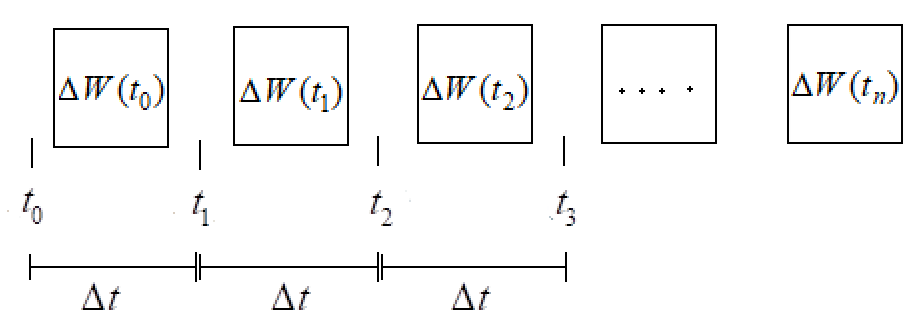
**General Stochastic Processes**

First we’ll do general stochastic processes, and then in next file I’ll simplify to Wiener processes. I’ll be including simplifications afforded if the processes has a Haar measure, as discussed in the RMT files.

**Discontinuous Stochastic Processes**

Consider a sequence of random, independent, identically distributed, possibly vector, variables, Δ**W**(tm), spaced a ‘distance’ Δt apart,



And let’s say they have a probability distribution PΔt(Δ**W**(tm)), which we may formally take to depend on Δt. The distribution is often taken to be normal,



Sometimes it’s taken to be uniform. But we won’t commit ourselves to anything yet. I think we *may* have to consider Δ**W**(tm) to be a continuous random variable, but I may be wrong – look more into it. In any event, we can think of these variables as something like random impulses.

**Recursion relation for X(t­m) random variables**

Now let Δ**W**(tm) drive the fluctuations in another sequence of vector random variables **X**(tm) – might take these to be momenta.

Diagram

Description automatically generated

Two possibilities are:



And let these variables, which define a **stochastic process**, obey the recursion relation.



(this f not necessarily same as f in example above – though *could* be) This is so far a completely general recursion relation. So in principle, the probability distribution of a term in a Stochastic process depends on all ΔW preceding, or, equivalently, all X’s preceding. We are ultimately interested in working out the probability distribution of the random variable X(tn). First we’ll discuss some simplifying properties we’ll resume our stochastic process to satisfy.

A Markov process obeys the following, namely that the random variable only depends on the previous variable, not the whole sequence prior, i.e., ΔX(tn) = g(X(tn), ΔW(tn)). So technically, the illustration I have above most closely resembles a Markov process since it insinuates that X(t3), say, depends only on X(t2) and ΔW(t2). If we thought of ΔW(tn) as a random velocity, and X(tn) as position, then X(tn) would be a Markov process, because it would only depend on the prior X(tn-1) and ΔW(tn-1). In such a circumstance, we may say the marginal probability distribution of **X**(tn) has the following property:

A **Markov** process has the following property: 

And then we have another salient property to mention:

A **Martingale** has further property that: 

Both of those X(tm) examples in the brackets would be both Markovian and Martingales. Let’s say that our process is Markovian, basically, but allowing explicit time-dependence, interval dependence, like this:



So Δ**X**(tn) depends only on **X**(tn) and Δ**W**(tn). And **X**(tn+1) depends only on **X**(tn), and Δ**W**(tn). Given this, we may calculate the conditional probability distribution of Δ**X**(tn). This is given below, and note that **X**(tm), tm, and Δt will appear as free parameters:



Also should mention that Δ**W** may be complex valued, and so the measure would have to include both real/imaginary parts. The probability distribution of X(tm) itself would be given by:



**Recursion Relations for pdf of X(tn)**

Now that we have, in principle, the probability distribution of the X(tm)’s, we can calculate averages of any function of these variables. So the average of any function F(X(tm)) can be written in many different ways. The second line is in terms of all ΔW(tj≤m). The third line is in terms of just X(tm) and ΔW(tm) (which we can do since F depends only on X(tm) and ΔX(tm) (which itself depends on X(tm) and ΔW(tm))). And the fourth line is in terms of X(tm) and ΔX(tm).



Note the order of the last expression. We have to write it this way because PΔt(ΔX) has X in it as a free parameter of sorts; so we have to do that integral first. It might seem odd that a probability distribution function, which depends on many ΔW variables, can be written in terms of a single variable X(tm). But just remember that we can obviously do it for two variables, and in any event, the d.o.f. of X are just (-∞, ∞), and this is the same regardless whether it’s a function of 1 or ∞ variables. In particular, we can write an evolution equation for the probability distribution function by letting F(**X**(tm+1)) = δ(**ξ**(tm+1) – **X**(tm+1)). And a recursion relation for P(X(tm)) follows as:



(really the argument of P would be ξ(tm+1), but I’m switching it back to X(tm+1) because it flows better) We can do the ΔW(tm) integral first, which would require a Jacobian, but I’ll ignore that and just formally write it as:



Or we can do it another way; we can perform the dX(tm) integral first. Parenthetically, remember X(tm+1) is just a spectator/independent variable here, and ΔX(tm) will depend on X(tm) probably, as well as ΔW(tm). Since X(tm) appears within ΔX(tm), to do the dX(tm) integral we’ll need a Jacobian to employ the delta function. Let JX(tm+1)X(tm) = |∂X(tm+1)/∂X(tm)|, and we’ll have this:



In the second line I’m trying to say that now we’ve solved for **X**(tm) in terms of **X**(tm+1) and Δ**W**(tm) in the delta function. And the Jaobian is what we get for that change of variable in the delta function. In the third line I’m just changing notation. And in the last line, I’m assuming ΔX(tm) is now evaluated in terms of X(tm+1) and ΔW(tm). Since X(tm+1) = X(tm) – ΔX(tm), by the δ function, we can write it in a slightly more suggestive form, which lends itself more readily to perturbative expansion:



**Simplification of pdf for X(tn) if X(tn) has Haar measure**

We’d like to get rid of the Jacobian, and this can be done by taking advantage of the matrix space that the variables X(t) are embedded in.



Then we can write the probability distribution function of the X variables as:



where dμ(**X**(tm)) = μ(**X**(tm))d**X**(tm) is the ‘volume’ in matrix space corresponding to the ‘volume’ d**X**(tm) in coordinate space. Then backing up a little to a previous expression, we can write,



Finally using the Haar measure property (see Random Matrices Measure

file):



we obtain:



which we can write as, since it’s convenient for perturbative purposes:



This integro-difference equation would appear to be a simpler starting point than the former ones. And both these integro-difference equations may be converted to purely integral equations with application of a z-transform, for instance. Still, these equations are usually prohibitively difficult to solve. Going to the continuum limit drastically simplifies things, though at the cost of not fully capturing the behavior for large fluctuations.

**Continuous Stochastic Processes**

Now let’s consider continuous processes. These could be obtained by adding more points within the sequence (and thereby adding random variables), while yet shrinking their std, say. A process would be continuous if:



Well, shouldn’t this be a δ function rather? Something like P(**X**(t)) = δ(**X**(t)). I think so. Anyway, it would seem to follow from this that we can write:



An example would be a Ornstein-Uhlenbeck process I think.



A process is **independent** if X(b)-X(a) is independent of X(b’) – X(a’) as long as the intervals don’t overlap. A process is **homogeneous** if X(b)-X(a) = X(b-a). A process fulfilling all of these stipulations would be a **Levy** **process**. A **Wiener process** is one such, and is given by the distribution:



**Making transition from discrete to continuous**

We can approximate our discrete stochastic process as a continuous one, though at the cost of getting the large fluctuations a bit wrong. When we do the same for harmonic oscillations in a material, we lose information about the eigenvalue spectrum for large k’s. But the eigenfunctions in the continuum limit look pretty much like those in the discrete case. The general procedure, it seems, is to artificially interpolate our process with more points in between, separated by dt. And we will call the new incremental driving random variables d**W**(tm).



Then we must decide on the probability distribution Pdt(d**W**(t)) of these new variables. In the continuum limit, we will typically need only the first two moments of the distribution. Well we only need the distribution out to order dt, and if we presume P(dW), whatever it is, to decay fast enough, then should have that we only need go out to order (dW)2. Or said another way. Since the pdf of the continuous process increment dX (or dW) goes to a delta function as dt → 0, the expectation of the moments must go to zero as dt → 0 as well. Presumably they will go as (dt)β where β is any power. But higher moments will go as higher powers than lower ones and so there is an upper bound to how high a moment we need (we can stop when β = 1). Typically this will be the second moment.

Physical intuition helps. Often the Δ**W**(tm) represent displacement, or force, or in our case, potential fluctuations of some sort. In the displacement example, then we might say the following, and match average and variance.



But if it represents force, then we’d want the impulse over the time Δt to be preserved, and so we’d say this (assuming zero average impulse):



Where R2 and F2 are tensors. Let’s introduce a Brownian motion increment, defined by:



and since this is as far as we’ll need to go, moment wise, we can pretend that these are all the moments, and therefore that d**W** is normally distributed. So would could just say for example, that



But I’ll not presume this to be the case until we get to Weiner Processes.

**Recursion relation for X(t)**

Then we can rescale our dW as necessary, and write, for some function g, not necessarily the same as the one above, that:



We’ll note that d**W**B(t) is in a sense on the order of √dt. And expanding dX out to order dt, assuming it’s well-defined, we’d have something like the following. There are two possible contributions to O(dt) term – one being dt, and the other d**W**Bd**W**B.



Another, maybe more heuristic possibility, often taken in the context of studying these GDMPK equation things, is to simply ‘Taylor expand’ the relevant variables in a stochastic series of the form above, and then postulate the a and b terms. This can be dicey as sometimes a = 0, like with the white noise model, even when you wouldn’t necessarily think it to be. Finally, as is done in Mello’s DMPK work, one might use a max entropy hypothesis to guess the form of p(dX) [not P(dX) per seʹ – you’d need to know the measure for *that*].

**Recursion relation for pdf of X(t)**

Anyway, might want to note, for concreteness’ sake, that the probability distribution corresponding to the dX above would be something like (assuming our boxed pdf for dW):



So we have:



But I’ll not presume this to be the case until we get to Weiner processes. Now reprising our discrete discussion, we can say, for any function F(X(t)), that it’s average is given by:



Expanding out to first order in dt, and dividing by dt, we get the following:



Assuming the small dt limit exists, we have a differential equation for the evolution of <F(**X**(t))>. Such an equation can sometimes be solved self-consistently. But if not, a Taylor series expansion could be developed by repeatedly applying the time-development operator.



Moving on to the continuum analogues of our discrete recursive relations for the probability distribution, we’d have:



And expanding the δ out in powers of dX(t) up to O(dt), renaming X(t+dt) as ξ, we have:



Integrating by parts (or alternatively, just switching the derivative ∂/∂**X**(t) to ∂/∂**ξ**), where all X(t) within dX(t) would be evaluated at ξ,



Dividing by dt, we have:



An alternative is to use the other discrete formula where the X(t) average was formally done first.



To develop a differential equation, the Jacobian would have to be expanded in powers of dX(t) as well, which would be dispreffered.

**Simplification of pdf for X(t) if X(t) has Haar measure**

Better would be to go to the other formula, and make the expansion.



Renaming X(t+dt) as ξ again, and expanding in powers dX(t), we get the following. Note that *this* dX is evaluated in terms of X(t+dt) and dW.



**Characteristic Function**

Instead of considering the evolution of <δ(ξ - X(t))>, to get the probability distribution function P(ξ,t), it might be advantageous to consider the evolution of <exp(iυX(t))>, to get the characteristic function, basically a *moment generating function*.



The evolution equation of the characteristic function can be obtained from a Fourier transform of the FP equation. I’ll assume X’s evolution is something like dX/dt = ai + bijwj.



We can write it as an integral equation by introducing the inverse Fourier transform of the a’s, b’s, and P’s.



And,



And so,



So this is the evolution equation for the characteristic function. As we can see, it is written as a convolution – basically an integral equation. Can other functions’ evolution equations be written self-consistently? Now we kind of assumed that the Fourier transform of a and b existed. But if not; if for instance a, b = x, x2 or something, we can still use this procedure. We’ll just note that:



So this is a possible alternative. And likewise we’ll note that however we get an evolution equation for χ(**υ**), we may invert via Fourier transform somehow to get an evolution equation for p(**ξ**).

Let’s explore an alternative path. If the characteristic function is known, then the evolution of all moments of X are known. For instance, say X is an n-element vector, then for any sj, we’d have:



Conversely, if the evolution of all moments is known, then the evolution of the characteristic function can be obtained. Suppose we knew that:



where the action of the operators is:



Then consider:



Now note:



So we can write the time-development equation for χ(**υ**,t) as:



Then, if the probability distribution function equation is desired, an inverse Fourier transform can be taken of both sides. The dependence of the results on λ isn’t as conspicuous here, but is implicit within the evolution equations of the moments.



after which we get:



**Changing Variables**

There is a useful chain rule identity of note. Suppose we have some function F(Z(X)), which depends on X through some auxiliary variable, or set of auxiliary variables Z(X). And suppose that the evolution operator H(X) can be written as:



Then we can write H’s action on F as:



And so if the action of H(Z) and H(Z2) returns some function of Z alone, then we can write a self-consistent equation for the probability distribution, or characteristic function of Z.