**Time Series 1**

Here’s some solved examples, to aid with intuition and such. Note we’ll repeatedly use:



So proceeding,

**White Noise**

This is the simplest, boringest.



where ΔWn is a white noise variable. The white noise variable’s statistical properties are:



where DΔt is some arbitrary whatever number. And we can think of the series elements being separated by some standard time, Δt (I’ll take to be 1). What is ACF(k)?



So this series is stationary at least as far as ACF, because it only depends on k. And the autocorrelation function makes clear that points are only correlated with themselves. The correlation between xj and xj-k is only 1 if k is 0. So whether xj is above/below its mean has no bearing on if xj-k is. The probability distribution of xn would be a normal distribution with mean 0 and variance DΔt:



The conditional probability distribution of xn would be:



Since xn doesn’t depend on the other variables. The joint probability distribution would be:



Being a product of normal distributions, the joint distribution is also a normal distribution, with a given mean column matrix μ, and covariance matrix, σ. We already know what these are. So we can write:



**Random Walk**

A random walk is given by the following recursion relation:



We can solve this equation:



So,



What are some expectations?



and,



So,



The variance is:



What is ACF?



So this series is not stationary, as ACF depends on both m and n, not just m-n. We can see the correlator and variance grows with time (well, index), though the average does not. The correlation between xj and xj-k is nearly 1 if k is small. This means that if xj is above/below its mean (0), then xj-k will be too most likely. But this is less and less likely the larger the lag, k, is. The probability distribution of xn would be:



i.e., a normal distribution with mean 0 and variance nDΔt. What about the conditional distribution? Since xn = xn-1 + ΔWn, it should be given by a Normal distribution with mean xn-1 and variance DΔt, the variance of ΔWn. So,



What about the joint probability distribution?



Being a product of normal distributions, the joint distribution is also a normal distribution, with a given mean column matrix μ, and covariance matrix, σ. We already know what these are. So we can write:



*Going to continuum?*

We can turn our discrete difference equation into a stochastic differential equation by shrinking the time-interval, while keeping the correlation over Δt constant. Then we’d have:



and then taking continuum limit,



and finally,



where w(t) = dW/dt, and,



Can solve for X(t). It’s just:



which compares well to the exact solution. Note the correlation:



What about continuous random walk ACF ?



So there.

**Random Walk w/ non-white noise**

Now consider a random walk,



But with ΔWn given by some other distribution. Let’s say it’s binomial, with probability p of going up 1 and probability q = 1 – p of going down one, each multiplied by D. We can solve this the same way as above:



What are some expectations?



The conditional probability distribution would be approximately normal with a mean of xn-1 + Dp and a variance of Dpq.



**AR(1) model**

Consider the SDE,



where wn is a white noise variable. φ and β are slope and drift rate terms. Again, Δt is the time between series terms, and is just a constant, and we can set it to 1, but I’m leaving it to make going to the continuum limit more transparent. We can solve the equation (assuming you can as a deterministic series),



Need to multiply by an ‘integrating factor’ to turn the LHS into a pure difference. So let the factor be In. Then we have:



We need the LHS to be a pure difference. But we’ll have to work out what it’s the difference of. Assume its some (gnxn), where gn is to be determined.



These imply,



Equating the two yields,



So then we can say,



So our solution is:



so,



First couple terms are:



What is average? We clearly have:



The covariance matrix is:



So,



And the variance is:



What is ACF?



So this, also isn’t generally a stationary series, as it depends on j and k. Last thing, the probability distribution of xn would be:



What about the conditional distribution? Since xn = φxn-1 + βΔt + ΔWn, it should be given by a Normal distribution with mean xn-1 + βΔt and variance DΔt, the variance of ΔWn. So,



What about the joint probability distribution?



Being a product of normal distributions, the joint distribution is also a normal distribution, with a given mean column matrix μ, and covariance matrix, σ. We already know what these are. So we can write:



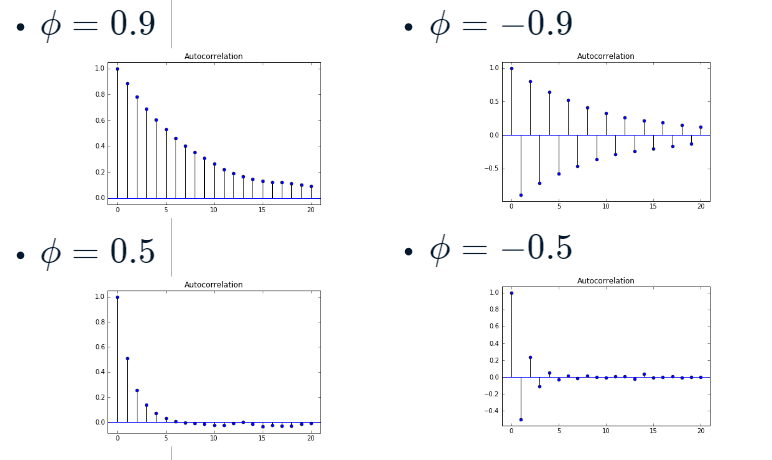
Let’s examine a few limits. What is the limit if |φ| < 1? Then we have:



So fluctuations from the far past will be severely damped. Those nearer the present will be still relevant. But one suspects that as n increases the total fluctuations won’t increase with time, as the window of relevancy should be constant, I’d say. And this is borne out by the <xn>2 calculation. Let’s look at ACFj(k) in the limit that both xj and xj-k are far down the series, so j, and j-k are large. If we have |φ| < 1, then



Remember the correlator tells us how jointly xj and xj-k vary. If correlator is 1, then if xj is above/below its mean, xj-k will be above/below *its* mean. And if correlator is -1, then if xj is above/below its mean, xj-k will be below/above *its* mean. And if correlator is 0, then there is no strong relationship. Well, i*n this limit*, we can see that the series *is* stationary, at least as far as its correlator, as its correlator is a function only of k. We also see that the correlations exponentially decay. And the smaller φ is, the more rapidly the correlations decay. Furthermore, if φ < 0, then they alternate sign too. This makes sense. If φ is positive, then the series has momentum and the only thing that would make it reverse momentum is an unusually large random noise element ΔWn kicking it in the other direction. So xm should be positive correlated with xm-k. But due to white noise, it should be less positively correlated with its predecessors the further we go back in the series. So the correlator should decrease with k. If φ is negative, the correlations alternate sign, as xm should be negatively correlated with xm-1, which is itself negative correlated with xm-2 (and so xm will be positively correlated with it), etc. Here’s a picture ACF(k) (see below) for different φ’s,



This matches what we expect from difference equation. What if φ = 1? This the random walk again. To evaluate the limit, we can use L’Hospital’s rule,



So we see that the variance grows with time (n). But the std = σ = √(nDΔt), which doesn’t grow as fast as μn = <xn>. So that’s good. As long as σn/μn << 1 maybe a regression or RNN will still work? Last, if we have large φ >> 1, then:



The std in the |φ| >> 1 case is σn = φ(n-1)DΔt. The average is μn ~ φn. So σ is on the order of μn – maybe a bit smaller. But it doesn’t look like it when you plot it because it’s exponential growth – always going up. Still, if you list the values, you can see that the uncertainty in xn is on the order of its average. So this just illustrates that noise could be quite present even if it isn’t super-obvious because it’s not fluctuating up and down. The fact that σn ~ μn would seem to make regression/RNN more problematic.

*Going to continuum?*

How does everything compare if we continuumize this series? Then we’d have:



Dividing by Δt, and then taking the limit Δt -> dt, we have:



What is the solution? I think the Ito and Stratonovich versions are the same. In this case, we can just integrate like a normal ODE. Solution should be:



And continuing,



So in the continuum limit we have:



Yep.

**AR(2) model**

Would look like this:



etc. As long as the φ’s are constants, it should be possible to analytically solve for xn, as it would be possible in the continuum case. We’ll separate this into a homogeneous and inhomogeneous solution. Let’s combine the two inhomogeneities into one by writing:



And we’ll separate the solution into a homogeneous solution and a particular solution.



For the homogeneous part, we’ll use the ansatz: xn(h) = 1/λn. Plugging this in, we have:



Here we can see how if the roots of the equation Φp(λ) = 1 – φ1λ – φ2λ2 lie outside the unit circle, then the series will be stationary. The roots are:



So the homogeneous solution will be:



Now we need to solve the initial conditions:



The solution to these equations is:



So our homogeneous solution is:



Now we need the particular solution, with initial conditions xn=0(p) = xn=1(p) = 0. Here we’ll employ the Green’s function technique. We’ll write:



Summation doesn’t include m = 0, 1 because DE really only is operative for n ≥ 2. We want this to satisfy the DE with the inhomogeneity. Plugging it in,



which requires:



Now we’ll look to solve our equation. Well, we already know the solution for n≠m. Repeating our work from above, this is:



(I’m including n = m in the top case, anticipating the fact that the formula for m < n will naturally give us the m = n result – see Difference Equations file). Let’s do initital conditions. These are:



This requires a> and b> to be zero I’d say. We can work out a<, b< by looking at different n’s. Let n = m . Then,



Now let’s do n = m + 1.



So our two equations are:



which have solutions:



So our GF is:



And our particular solution is:



And our complete solution is:



Well, let’s fill in fm = βΔt + ΔWn,



I guess it’s not getting any better, so:



where we recall,



and fn = βΔt + ΔWn. Pretty crazy looking. I haven’t tested this, so treat the solution with a grain of salt. But anyway. If modulus of a root is less than 1, then we get exponential growth. If moduli of both roots are 1, then we get power law behavior I guess. If moduli of both roots are greater than 1, then we get exponential decay. If the roots have complex parts, then we also get oscillatory behavior, and only oscillatory behavior if there is no real part to the roots. This seems true for both average and variance. What are the mean and variance of our variable? Well, the mean is obviously:



The covariance is:



Yummy. I’ll just stop here. So,



The variance would be the covariance with m = n:



which simplifies an incy bit to:



The probability distribution for xn ought to be:



for short. What about the conditional distribution? Since xn = φ1xn-1 + φ2xn-2 + βΔt + ΔWn, it should be given by a Normal distribution with mean φ1xn-1 + φ2xn-2 + βΔt and variance DΔt, the variance of ΔWn. So,



What about the joint probability distribution?



And again, as this is a product of normal distributions, it will itself be a normal distribution with a given mean and covariance matrix. So it will look like this (not enough space to write these out in the box):

