Week 7
Diffusion processes

Jonathan Goodman

October 28, 2013

1 Introduction to the material for the week

This week we discuss a random process X_t that is a *diffusion process*. A diffusion is a continuous time Markov process with continuous sample paths. Brownian motion is a diffusion process. The Poisson arrival process, N_t = number of arrivals up to time t, is not a diffusion because its sample paths are not continuous – N_t is discontinuous at any arrival time. The process $X_t = W_t^2$ is a diffusion because X_t is a continuous function of t and it is Markov. The process X_t below is a continuous function of t, but it is not a Markov process. It is an iterated integral involving Brownian motion, written in two equivalent ways:

$$f_s = \int_0^{t_1} W_r \, dr$$
$$X_t = \int_0^t f_s \, dW_s$$
$$X_t = \int_0^t \left(\int_0^s W_r \, dr \right) dW_s \, .$$

A more or less equivalent definition, and the more important one for applications, is that a diffusion is a process that satisfies a *stochastic differential equation*. An ordinary differential equation specifies the rate of change of X_t as a function of X_t and t. We might call the solution an *ordinary process* (nobody does). A stochastic differential equation for a diffusion process specifies the *infinitesimal mean*, or *drift*, and the *infinitesimal variance*, or *noise*. A diffusion process with drift $a(x,t)$ satisfies

$$\mathrm{E}[\, \Delta X_t \mid \mathcal{F}_t] = a(X_t, t)\Delta t + O(\Delta t^2) \, . \tag{1}$$

Here, $\Delta t > 0$, and $\Delta X = X_{t+\Delta t} - X_t$ is the forward looking change. The differential version for $dX_t = X_{t+dt} - X_t$ is

$$\mathrm{E}[\, dX_t \mid \mathcal{F}_t] = a(X_t, t)dt \, . \tag{2}$$

This is a little more than an informal restatement of (1). Last week, we said that the definition of differential is that you know what you get when you integrate it. For (2), this means that

$$\mathrm{E}[\, X_t] = X_0 + \int_0^t \mathrm{E}[\, a(X_s, s)]\; ds\;.$$

If X_t is a one dimensional process, the infinitesimal variance, $\mu(x,t)$, is defined by

$$\mathrm{E}\big[\, \Delta X_t^2 \mid \mathcal{F}_t\big] = \mu(X_t, t)\Delta t + O(\Delta t^2)\;. \tag{3}$$

The differential version of this is

$$\mathrm{E}\Big[\, (dX_t)^2 \mid \mathcal{F}_t\Big] = \mu(X_t, t)dt\;. \tag{4}$$

The integral relating to this is the *quadratic variation*.

$$\int_0^t (dX_t)^2 = \int_0^t \mu(X_s, s)\, ds\;. \tag{5}$$

As in assignment 4, the left side is defined as the limit

$$\lim_{m\to\infty} \sum_{t_j<t} (\Delta X_j)^2\;. \tag{6}$$

A theorem in the spirit of Ito's lemma says the limit exists and is equal to the right side.

An "ordinary process", one without noise, has a differential update formula $dX_t = a(X_t, t)dt$. A diffusion process differential update formula would need to take the form

$$dX_t = a(X_t, t)dt + (\text{random mean zero variance } \mu(X_t, t)dt)\;.$$

The last term needs to have mean zero and variance $\mu(X_t, t)dt$ even conditional on $\mathcal{F}_t$. This may be achieved, and this is not the only way, by taking $b(x,t) = \sqrt{\mu(x,t)}$ and

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t\;. \tag{7}$$

The noise term $b(X_t, t)dW_t$ meets our requirements. It has mean zero and variance $b(X_t, t)^2 dt = \mu(X_t, t)dt$, even conditioned on knowing $\mathcal{F}_t$. The expression (7) is an Ito *stochastic differential equation*, or *SDE*. We have argued informally that any diffusion process can be described as being the solution of an SDE.

Most interesting "ordinary" processes are multi-dimensional. This is true also for diffusion processes. The state X_t can have n components

$$X_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{n,t} \end{pmatrix}\;.$$

The definition of drift may be written as (1) or (2), except now the quantities on both sides are n component vectors. The infinitesimal variance is the $n \times n$ matrix that describes the covariance matrix of dX_t

$$\mathrm{E}\Big[\left(dX_t\right) \left(dX_t\right)^t \mid \mathcal{F}_t \Big] = \mu(X_t, t)dt \; . \tag{8}$$

An SDE of the form (7) can describe multi-component, or *multivariate*, diffusions. The drift, $a(x,t)$, is an n component vector. The coefficient $b(x,t)$ is a matrix with n rows so that $b(x,t)b^t(x,t) = \mu(x,t)$. We discuss this below.

The goal for this week is to develop much of the stochastic calculus for general diffusions. This means finding general diffusion versions of the Ito integral, some backward equations, Ito's lemma, and quadratic variation. We seem to have a lot to do, but the derivations and proofs should start to seem routine.

2 Some diffusion processes

Here are some diffusion process models:

Geometric Brownian motion. A *geometric Brownian motion* with *growth rate* μ and *volatility* σ, is a process that satisfies the SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t \; . \tag{9}$$

The *Black Scholes* theory of stock option pricing uses geometric Brownian motion to model the price at time t of a stock. Ito's lemma from last week allows us to check that the solution is

$$S_t = S_0 e^{\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t} \; . \tag{10}$$

This week's version of Ito's lemma allows you to derive this formula in a different and possibly more natural way.

Stochastic volatility model. The volatility parameter σ determines the amount of noise in a geometric Brownian motion. It is observed that the level of noise in stock price processes seems to change unpredictably with time. A *stochastic volatility* model replaces the constant σ with the stochastic process σ_t, which is governed by its own SDE. One model of this type is expressed by the SDE system

$$dS_t = \mu S_t dt + \sigma_t S_t dW_{1,t} \tag{11}$$

$$d\sigma_t = \lambda \left(\overline{\sigma} - \sigma\right) dt + \nu dW_{2,t} \; . \tag{12}$$

The volatility process is a linear *mean reverting* process with equilibrium volatility $\overline{\sigma}$, mean return rate λ, and volatility of volatility (or *volvol*) ν.

Linear Gaussian models. A linear Gaussian continuous time model has the form

$$dX_t = AX_t dt + BdW_t \ . \tag{13}$$

We said that X_t has n components, which forces A to be an $n \times n$ matrix. But W could have n or fewer components, so $m \leq n$ will be the number of components of W:

$$W_t = \begin{pmatrix} W_{1,t} \\ W_{1,t} \\ \vdots \\ W_{m,t} \end{pmatrix} \ .$$

We assume here that distinct components $W_{j,t}$ and $W_{k,t}$ are independent, which is expressed by the formula

$$\mathrm{E}\big[\, dW \, dW^t \mid \mathcal{F}_t \big] = I_{m\times m} \, dt \ . \tag{14}$$

Each component of W is a *source of noise*. The informal *Ito rule* is equation without the expectation on the left.

Here is a simple example that came up for me modeling surface oscillations of stars. Suppose the surface displacement is Y_t. A simple harmonic oscillator model is

$$\ddot{Y} = -\omega_0^2 Y \ .$$

This has solutions $Y_t = A\cos(\omega_0 t) + B\sin(\omega_0 t)$. You can add a friction "force" with friction coefficient γ, which gives

$$\ddot{Y} = -\omega_0^2 Y - \gamma \dot{Y} \ . \tag{15}$$

If γ and ω_0 are positive, and γ is small relative to ω, then Y_t will be an exponentially decaying oscillation. But real stars have damped oscillators that are "driven" by random forces, presumably from turbulence within the star. We would like

$$\ddot{Y} = -\omega_0^2 Y - \gamma \dot{Y} + \sigma(\text{random force}) \ .$$

We formulate this in the general form (13) using the standard idea from differential equations: introduce the velocity variable $V_t = \dot{Y}_t$. The single second order equation (15) is equivalent to the following system of two first order equations:

$$\begin{aligned} \dot{Y}_t &= V_t \\ \dot{V}_t &= -\omega_0^2 Y_t - \gamma V_t \ . \end{aligned}$$

The random forcing for the surface oscillation model should go in the second equation only, because the first equation is an exact relation that defines V. We write a system of two Ito style differential equations for the noisy oscillator model

$$\begin{aligned} dY_t &= V_t \, dt \\ dV_t &= \big(-\omega_0^2 Y_t - \gamma V_t\big) \, dt + b \, dW_t \ . \end{aligned}$$

In matrix form, this is

$$d\begin{pmatrix} Y_t \\ V_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix} \begin{pmatrix} Y_t \\ V_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ b \end{pmatrix} dW_t \; .$$

This is a model of the form (13) with $n = 2$, $m = 1$, $A = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix}$, and $B = \begin{pmatrix} 0 \\ b \end{pmatrix}$.

3 Continuous and discontinuous path models

A process can satisfy (1) and (4) without being a diffusion or even being continuous. You can tell a *jump process* (a process with discontinuities) from a diffusion by looking at moments beyond the variance. A single component diffusion should satisfy

$$\mathrm{E}\Big[\, (\Delta X_t)^4 \mid \mathcal{F}_t \Big] \leq C\Delta t^2 \; . \tag{16}$$

This is consistent with the idea that $\Delta X \sim \sqrt{\Delta t}$, like Brownian motion. It is a theorem that Markov processes that satisfy this fourth moment condition have continuous sample paths (almost surely). We will not even sketch the proof. Brownian motion has $\mathrm{E}\Big[\, (\Delta W_t)^4 \mid \mathcal{F}_t \Big] = 3\Delta t^2$. Jump processes have

$$\mathrm{E}\Big[\, (\Delta X_t)^4 \mid \mathcal{F}_t \Big] \approx C\Delta t \; . \tag{17}$$

The *Poisson arrival process* with *arrival rate*, or *intensity*, λ is the simplest jump process. In this model, there is an *arrival* in time interval $(t, t + dt)$ with probability λdt. All arrivals are independent. The process is a function N_t, which is equal to the number of arrivals in the interval $[0, t]$. This is an increasing function of t that jumps discontinuously at each arrival time. The increments, and their probabilities, are

$$\Delta N_t = N_{t+\Delta t} - N_t = \begin{cases} 0 \; , & \mathrm{P}(\Delta N = 0 \mid \mathcal{F}_t) \approx 1 - \lambda\Delta t & \text{(no arrivals)} \\ 1 \; , & \mathrm{P}(\Delta N = 1 \mid \mathcal{F}_t) \approx \lambda\Delta t & \text{(one arrival)} \\ \geq 2 \; , & \mathrm{P}(\Delta N \geq 2 \mid \mathcal{F}_t) = O(\Delta t^2) & \text{(more than one, unlikely).} \end{cases}$$

We write $\mathrm{P}(\Delta N = 1) \approx \lambda\Delta t$ because the interval Δt is not infinitely small. An exact calculation reaches the same conclusions we are about to get. These probabilities give

$$\begin{aligned} \mathrm{E}[\, \Delta N \mid \mathcal{F}_t] \; &= \; 0 \cdot \mathrm{P}(\Delta N = 0 \mid \mathcal{F}_t) + 1 \cdot \mathrm{P}(\Delta N = 1 \mid \mathcal{F}_t) + O(\Delta t^2) \\ &\approx \; \lambda\Delta t \; . \end{aligned}$$

This turns out to be exact, as the various approximation errors cancel. The expected number of arrivals in an interval is exactly equal to rate multiplied

by the length of the interval, which gives $\lambda\Delta t$ here. The higher moments are approximately the same, though not exactly. The calculation above gives

$$\mathrm{E}[\,(\Delta N)^p \mid \mathcal{F}_t] = \lambda\Delta t + O(\Delta t^2)\ ,$$

for any $p > 0$.

The Poisson process shows why the higher moments are important. If you look just at the first and second moments, which is (1) and (3) in this case, the Poisson process looks like a diffusion with $a = 1$ and $\mu = 1$. But the Poisson process in not a diffusion. We see this from the sample paths, which are discontinuous. We see this in the analysis, as the Poisson process fails to satisfy the fourth moment condition (16). The fourth moment of a diffusion process is much smaller, smaller by a factor of Δt, than the fourth moment of ΔN.

3.1 Fat tailed distributions

This analysis of the Poisson process illustrates the phenomenon of *fat tailed* distributions. The *tails* of a probability density $u(x)$ are the parts of the graph for large $|x|$. You can think of the "bell-shaped" Gaussian graph as a picture of a mouse with no head and two tails. The size of the tails is the rate at which $u(x)$ goes to zero as $x \to \pm\infty$. Exponential decay, $u(x) \leq Ce^{-a|x|}$, is a "thin" tail. Power law decay with exponent q, $u(x) \sim C\,|x|^{-q}$ for large $|x|$, is "fat" tails. The tails are fatter when q is smaller. We cannot have $q \leq 1$ because

$$\int_{-\infty}^{\infty} u(x)\,dx < \infty\ .$$

For example, if $u(x) \sim C/x$ as $x \to \infty$, then $\int_a^\infty u(x)dx = \infty$, which is impossible for a probability density.

You can tell a fat tailed distribution by its large or infinite moments. If X is a random variable with a fat tail distribution, then $\mathrm{E}[\,|X|^p] = \infty$ if $p > q-1$. The Student t distribution with n degrees of freedom (see Assignment 1) has density

$$u_n(x) = \frac{C_n}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}\ .$$

is a fat tailed distribution with $q = n+1$. If X is a Student t random variable with $n = 3$ degrees of freedom, then $\mathrm{E}\left[\,X^2\right] = 3 < \infty$ but $\mathrm{E}\left[\,X^4\right] = \infty$. (It is easy to see that $\mathrm{E}\left[\,X^2\right] < \infty$ and $\mathrm{E}\left[\,X^4\right] = \infty$. Computing $\mathrm{E}\left[\,X^2\right] = 3$ is a challenge.) We could try to estimate $\mathrm{E}\left[\,X^4\right]$ by saying that $\mathrm{E}\left[\,X^2\right] = 3$ implies that $|X| \sim \sqrt{3}$, so $\mathrm{E}\left[\,X^4\right] \sim \sqrt{3}^4 = 9$. Fat tails ruin this argument. The expected value is determined not by typical X values on the order of $\sqrt{3}$, but low probability X values that are much larger than $\sqrt{3}$. These large X values account for the fact that $\mathrm{E}\left[\,X^4\right] \gg \left(\mathrm{E}\left[\,X^2\right]\right) > 2$.

If you have a fat tailed distribution, you might not know which moments are finite. But there is one thing you can say: if $\mathrm{E}[\,|X|^{p_1}] < \infty$, and if $p_2 < p_1$, then

$\mathrm{E}[\,|X|^{p_2}] < \infty$. This is because larger p makes x^p larger for large x, which makes it harder for the integral of $x^p u(x)$ to be finite. *Jensen's inequality* confirms this intuition in a quantitative way we will use later. This concerns a convex function $\phi(y)$.

$$\mathrm{E}[\,\phi(Y)] \geq \phi\left(\mathrm{E}[\,Y]\right) \ . \tag{18}$$

For example, if $\phi(y) = y^2$, we get

$$\mathrm{E}\left[\,Y^2\right] \geq \left(\mathrm{E}[\,Y]\right)^2 \quad \Longrightarrow \quad \mathrm{E}[\,Y] \leq \sqrt{\mathrm{E}[\,Y^2]}\,.$$

This inequality was used weeks ago, and proven using the Cauchy Schwarz inequality. Jensen's inequality gives a different proof. Now suppose $Y \geq 0$ and take a power $p > 1$ so that $\phi(y) = y^p$ is a convex function of y. The following more general inequality follows as the previous one did:

$$\mathrm{E}[\,Y] \leq \left(\mathrm{E}[\,Y^p]\right)^{1/p} \ . \tag{19}$$

This says that if the p^{th} moment of Y is finite then the first moment is finite. The specific case $Y = \Delta X^2$ and $p = 2$ could have been done using Cauchy Schwarz:

$$\mathrm{E}\left[\,\Delta X^2\right] \leq \sqrt{\mathrm{E}[\,\Delta X^4]}\,. \tag{20}$$

The case $Y = |\Delta X|^3$ and $p = \frac{4}{3}$ is what we need in Section 4:

$$\mathrm{E}\Big[\,|\Delta X|^3\Big] \leq \left(\mathrm{E}\left[\,\Delta X^4\right]\right)^{3/4} \ . \tag{21}$$

The Poisson process follows this pattern. $\mathrm{E}[\,\Delta N] = \lambda \Delta t$ does not imply that $\Delta N \sim \lambda \Delta t$ is a typical outcome. On the contrary, the expectation is determined by the event $\Delta N = 1 \gg \lambda \Delta t$, which has probability $\lambda \Delta t$. The rare $\Delta N = 1$ events also determine the expectations of the higher moments.

It is crucial in many situations to model the tails of distributions correctly. Tail events, by definition, have low probability. But they can have a big impact on overall behavior. In the Poisson process, it is a tail event to have an arrival in a small interval $[t, t + dt]$. Yet, without arrivals there would be no process. People have gone broke using stock price models with exponentially thin tails. A Gaussian random variable 4σ event has tiny probability $\mathrm{P}(|X - \mu| > 4\sigma) = 6.4 \times 10^{-5}$, which is about one day every 200 years. A fat tailed distribution could have $\mathrm{P}(4 \text{ sigma}) = \frac{1}{16} = 6.25 \times 10^{-2}$, be Chebychev's inequality, which is roughly 1000 times more likely. Heavy tailed distributions arise in many fields – the sizes of earthquakes, the sizes of car insurance payouts, the sizes of internet server requests, the sizes of stars, the list goes on.

At this moment in stochastic calculus, the takeaway is that you can show that X_t is a diffusion process by showing that ΔX_t has light tails. That means that the fourth moment is on the order of the square of the second moment: $\mathrm{E}\Big[\,(\Delta X)^4\Big] \leq C\,\left(\mathrm{E}\Big[\,(\Delta X)^2\Big]\right)^2 \approx C\,(\mu \Delta t)^2$.

3.2 Poisson processes in more detail

We describe three different ways to simulate a simple Poisson arrival process. These are (1) direct simulation, (2) event driven simulation, (3) rain making. Each reveals something different about arrival processes. Each generalized to a different class of more complex simulation methods.

Direct simulation is a form of time stepping. You choose a time step Δt and get discrete times $t_j = j\Delta t$. The values $N_j \sim N_{t_j}$ as $N_{j+1} = N_j + \Delta N_j$. The increment is a Bernoulli random variable with $\mathrm{P}(\Delta N_j = 1) = p = \lambda \Delta t$, and $\mathrm{P}(\Delta N_j = 0) = 1 - p$. All ΔN_j values are independent. You "toss a coin" in the computer using a uniform random number generator. If U is uniformly distributed in $[0, 1]$, then $\mathrm{P}(U < p) = p$ (duh). In `R` you can use this as follows

```
p = lam*dt             # prob of an arrival in a dt interval
t = 0                  # initialize the time variable
N = 0                  # Set N_0 = 0 -- start with no arrivals
while ( t < T ) {            # time steps up to time T
   U  = runif( 1, 0., 1.)    # 1 uniform random variable in [0,1]
   dN = 0                    # the most likely value
   if ( U < p ) dN = 1       # generate an arrival
   N = N + dN                # add in the increment
   t = t + dt                # t_{j+1} = t_j + dt
  }
```

This method has many faults. For one thing, it is slow. It asks over and over "is there an arrival here?", most often with the answer "no". For another thing, it is not exact. The probability of an arrival is not exactly $\lambda \Delta t$, there is an $O(\lambda \Delta t^2)$ probability of more than one arrival in an interval of length Δt. We do not say where in $[t, t + \Delta t]$ the new arrival falls.

Event driven simulation simulates only the events that happen, rather than a large number of non-events. The events here are the arrivals. Suppose arrival k happens at the arrival time T_k. The *inter arrival* times are $S_k = T_k - T_{k-1}$. These are independent exponential random variables with rate constant λ. The probability density of S_k is $v(s) = \lambda e^{-\lambda s}$, if $s > 0$. You can generate an exponential random variable from a uniform using the formula $S = -\log(U)/\lambda$. It is convenient to suppose $T_0 = 0$, so the first arrival is at time $T_1 = T_0 + S_1 = S_1$, and so on. The arrival counting function, is $N_t = k$ for $t \in [T_{k-1}, T_k]$. You can wonder whether $N_{t_k} = k$ or $N_{t_k} = k + 1$. This *discrete event simulation strategy* is obviously faster than direct simulation.

The *rain making* approach first asks how many arrivals there were in the interval of interest $[0, T]$, which is N_T. The probabilities are $\mathrm{P}(N_T = n) = e^{-\lambda T} \frac{(\lambda T)^n}{n!}$. This is the *Poisson distribution* with parameter λT. The `R` statement `N = rpois( 1, lam*T)` generates a Poisson random variable. Once we know N_T, the arrival times may be generated using i.i.d. uniformly distributed random variables in $[0, T]$. In `R`, this could be a vector of length N containing N independent uniforms `Tv = runif( N, 0., T)`. The list `Tv` contains the arrival

times, but not in order. You can find the k^{th} arrival time by sorting the list: `sort(Tv)`. Then `Tv[k]` will be the k^{th} arrival time T_k.

There are good algorithms for generating the necessary Poisson N. I don't know how the `R` routine `rpois` works. You will learn how it might work if you take my Monte Carlo class. It is not complicated, but too complicated for this class.

The name rain comes from thinking of the first step as deciding how many raindrops there will be and the second step as letting them land uniformly in the interval $[0, T]$.

There are many generalizations of these methods. Unfortunately, many problems defy clever solutions and call for direct simulation. Event driven simulation can simulate discrete state space continuous time processes such as continuous time Markov chains. The rain algorithm can be used to generate multi-dimensional Poisson processes. Such a process is defined by having a "drop" in a small set A with approximate probability $\lambda \, \text{area}(A)$, with non-overlapping areas being independent.

4 Backward equations

We derive the backward equation for the value function $f(x,t) = \mathrm{E}_{x,t}[\, V(X_T)]$. We give a treatment that generalizes the direct treatment in Section 3.1 of Week 6. The new points this week include: (1) a more general PDE that incorporates the drift and noise coefficients, (2) a derivation that uses the fourth moment conditions more explicitly than last week, (3) a backward equation for multi-component diffusions, which is a PDE with more than one x variable.

As in Week 6, the value function has two equivalent derivations:

$$f(x,t) = \mathrm{E}_{x,t}[\, V(X_T)] \quad , \quad f(X_t, t) = \mathrm{E}[\, V(X_T) \mid \mathcal{F}_t] \ . \tag{22}$$

We use the tower property, as we did last week, to express $f(\cdot, t)$ in terms of $f(\cdot, t + \Delta t)$. We keep the notation $\Delta X = X_{t+\Delta t} - x$, so $X_{t+\Delta t} = x + \Delta X$, and

$$f(x,t) = \mathrm{E}_{x,t}[\, f(x + \Delta X, t + \Delta t)] \ . \tag{23}$$

We expand the right side to first order in Δt and second order in ΔX. Expressions are shortened by leaving out (x,t) arguments, as in $f(x,t) \rightsquigarrow f$. We continue the derivation assuming a one dimensional X_t. The Taylor calculations of last week give

$$\begin{aligned} 0 =& \partial_x f \, \mathrm{E}_{x,t}[\, \Delta X] + \frac{1}{2} \partial_x^2 f \, \mathrm{E}_{x,t}\Big[\, (\Delta X)^2 \Big] + \partial_t f \Delta t \\ &+ \mathrm{E}\Big[\, A \, |\Delta X|^3 \Big] + \mathrm{E}[\, B \, |\Delta X|] \, \Delta t + C \Delta t^2 \ . \end{aligned}$$

The terms on the second line are Taylor remainder terms, exactly as they were last week. The infinitesimal mean (1) and variance (3) determine the terms on

the top line. If X_t is a diffusion and satisfies the continuous path fourth moment condition (16), then the first remainder term on the second line satisfies

$$\begin{aligned} \mathrm{E}\Big[\, A\,|\Delta X|^3\Big] &\leq C\,\mathrm{E}\Big[\,|\Delta X|^3\Big] \\ &\leq C\,\left(\mathrm{E}\big[\,\Delta X^4\big]\right)^{3/4} \\ &\leq C\,\left(\Delta t^2\right)^{3/4} \\ &= C\Delta t^{3/2}\,. \end{aligned}$$

We used the inequality $\mathrm{E}\Big[\,|\Delta W|^3\Big] = C\Delta t^{3/2}$ in the same way last week. Now we "do the math", which means putting all this information into the Taylor expansion then collecting terms.

$$0 = \partial_x f\, a(x,t)\Delta t + A\Delta t^2 + \frac{1}{2}\partial_x^2 f\,\mu(x,t)\Delta t + B\Delta t^2 + \partial_t f\Delta t + C\Delta t^{3/2}\,.$$

The desired backward equation appears when we divide by Δt and let Δt go to zero, substitute back the (x,t) arguments, and rearrange a little:

$$0 = \partial_t f(x,t) + a(x,t)\partial_x f(x,t) + \frac{1}{2}\mu(x,t)\partial_x^2 f(x,t)\,. \tag{24}$$

Of course, the final condition is $f(x,T) = V(x)$.

Brownian motion. Standard Brownian motion with no drift and variance t satisfies the SDE $dX_t = dW_t$. This implies that $X_t = W_t$, provided $X_0 = W_0$. This SDE has $a = 0$ and $b = 1$, so $\mu = b^2 = 1$. The backward equation for this case is $0 = \partial_t f + \frac{1}{2}\partial_x^2 f$, which is the backward equation we had before for Brownian motion.

Linear mean reversion, Ornstein Uhlenbeck.

A solution of the SDE

$$dX_t = -\gamma X_t dt + \sigma dW_t \tag{25}$$

with $\gamma > 0$ is a *mean-reverting linear Ornstein Uhlenbeck* process. In a time dt, X_t has a mean "drift" toward the origin with a speed depending on X_t. It "reverts" to the mean, which is zero in this example. The noise σdW_t stops X_t from settling down to zero as $t \to \infty$. This equation has $a(x,t) = -\gamma x$, and $\mu = b^2 = \sigma^2$. The backward equation is

$$0 = \partial_t f - \gamma x\partial_x f + \frac{1}{2}\sigma^2\partial_x^2 f\,. \tag{26}$$

There is an explicit solution for final condition $V(x) = e^{-x^2/2}$. The mechanics of this solution, via the ansatz method, are presented first, then the "physical" interpretation. An *ansatz* is a guess at the form of the solution. The

guess contains unknown parameters or functions, which then are determined using the equation. We start with the ansatz

$$f(x,t) = A(t)e^{-B(t)x^2/2} \; . \tag{27}$$

Time derivatives are denoted with a dot, as $\partial_t A(t) = \dot{A}(t)$. The derivative calculations are (leaving out the exponents where possible):

$$\begin{aligned} A(t)e^{-B(t)x^2/2} &\xrightarrow{\partial_t} \dot{A}e^{-} + \frac{-1}{2}A\dot{B}x^2e^{-} \\ A(t)e^{-B(t)x^2/2} &\xrightarrow{\partial_x} -ABx\, e^{-B(t)x^2/2} \\ &\xrightarrow{\partial_x} -AB\, e^{-} + AB^2x^2\, e^{-} \; . \end{aligned}$$

We substitute these calculations into the PDE and cancel the common exponential factor, giving

$$0 = \dot{A} + \frac{-1}{2}A\dot{B}x^2 - \gamma x\left(-ABx\right) + \frac{\sigma^2}{2}\left(-AB + AB^2x^2\right) \; . \tag{28}$$

The PDE (26) is satisfied if this equation is satisfied. As a function of x, this equation is a quadratic with an x^2 term and a constant term. The quadratic equals zero if both coefficients are equal to zero. This reasoning reduces (28) to two equations

$$(x^2 \text{ terms}): \qquad 0 = \frac{-1}{2}A\dot{B} + \gamma AB + \frac{\sigma^2}{2}AB^2x^2 \tag{29}$$

$$(x^0 \text{ terms}): \qquad 0 = \dot{A} - \frac{\sigma^2}{2}AB \tag{30}$$

There are explicit solutions for these differential equations. The $O(x^2)$ equation may be solved for B first. The first step of the ODE book recipe for the solution is to calculate

$$\begin{aligned} \dot{B} &= 2\gamma B + \sigma^2 B^2 \\ \frac{dB}{dt} &= 2\gamma B + \sigma^2 B^2 \\ \frac{dB}{2\gamma B + \sigma^2 B^2} &= dt \\ \frac{dB}{B\left(2\gamma + \sigma^2 B\right)} &= dt \; . \end{aligned} \tag{31}$$

Step 2 is the partial fractions expansion

$$\frac{1}{B\left(2\gamma + \sigma^2 B\right)} = \frac{a}{B} + \frac{b}{2\gamma + \sigma^2 B} \; .$$

The solution comes from multiplying out the right side:

$$1 = a\left(2\gamma + \sigma^2 B\right) + bB \; .$$

It is

$$a = \frac{1}{2\gamma} \ , \quad b = \frac{-\sigma^2}{2\gamma} \ .$$

so

$$\frac{dB}{B\,(2\gamma + \sigma^2 B)} = \frac{1}{2\gamma}\,\frac{dB}{B} - \frac{\sigma^2}{2\gamma}\,\frac{dB}{2\gamma + \sigma^2 B} \ .$$

This integrates to

$$\frac{1}{2\gamma}\,\log B - \frac{1}{2\gamma}\log(2\gamma + \sigma^2 B) + Const = \frac{1}{2\gamma}\log\left(\frac{B}{2\gamma + \sigma^2 B}\right) + Const$$

Step 3 is to integrate both sides of (31), with a constant of integration C that changes from line to line

$$\begin{aligned}
\frac{1}{2\gamma}\log\left(\frac{B}{2\gamma + \sigma^2 B}\right) + C &= t \\
\frac{B}{2\gamma + \sigma^2 B} &= Ce^{2\gamma t} \\
B &= 2\gamma C e^{2\gamma t} + \sigma^2 B C e^{2\gamma t} \\
\left(1 - \sigma^2 C e^{2\gamma t}\right) B &= 2\gamma C e^{2\gamma t} \\
\left(Ce^{-2\gamma t} - \sigma^2\right) B &= 2\gamma \\
B(t) &= \frac{2\gamma}{Ce^{-2\gamma t} - \sigma^2} \ .
\end{aligned} \tag{32}$$

This algebra is easy, but it isn't quick.

The final condition $f(x,T) = V(x) = e^{-x^2/2}$ determines final conditions for $A(T)$ and $B(T)$. The conditions that make the ansatz (27) match V clearly are

$$A(T) = 1 \ , \quad B(T) = 1 \ . \tag{33}$$

The B final condition determines the C in (32), as follows:

$$\begin{aligned}
1 &= \frac{2\gamma}{Ce^{-2\gamma T} - \sigma^2} \\
Ce^{-2\gamma T} - \sigma^2 &= 2\gamma \\
C &= \left(2\gamma + \sigma^2\right) e^{2\gamma T} \ .
\end{aligned} \tag{34}$$

This value of C puts the B formula into its final form

$$B(t) = \frac{2\gamma}{(2\gamma + \sigma^2)\,e^{2\gamma(T-t)} - \sigma^2} \ . \tag{35}$$

We simplify the discussion by taking $t = 0$ and large T, so that

$$B(0) \approx \frac{2\gamma}{2\gamma + \sigma^2} e^{-2\gamma T} \ . \tag{36}$$

The lesson we get from this is that $B(0)$ is exponentially small for large T.

The explicit formula for A will appear after more calculations. We learn what the required calculations are from the A equation:

$$\begin{aligned}
\dot{A} &= \frac{\sigma^2}{2} AB \\
\frac{\dot{A}}{A} &= \frac{\sigma^2}{2} B \\
\frac{dA}{A} &= \frac{\sigma^2}{2} B\, dt \\
\ln(A) &= \frac{\sigma^2}{2} \int B\, dt + C \\
A &= C \exp\left(\frac{\sigma^2}{2} \int B\, dt \right) \ . \qquad (37)
\end{aligned}$$

The required calculation is the integral of (32), which then is exponentiated. It may seem too complicated at this point to be worth it. But in the real world, as opposed to the classroom, you have days or weeks to solve problems like this. The algebra is quick if it "only" takes a few hours. Furthermore, we saw already with the B approximation (36) that a complicated formula can contain simple information.

The simpler B formula (32) is a simpler starting point for integration. The first step is to undo an algebraic simplification[1] by multiplying by $e^{2\gamma t}$.

$$B(t) = \frac{2\gamma e^{2\gamma t}}{C_B - \sigma^2 e^{2\gamma t}} \ .$$

We write C_B for the constant from the B equation, and C_A for the A equation integration constant. The natural substitution is $u = e^{2\gamma t}$, which comes with $du = 2\gamma e^{2\gamma t} dt$. The integral is

$$\begin{aligned}
\int B(t) dt &= \int \frac{2\gamma e^{2\gamma t}}{C_B - \sigma^2 e^{2\gamma t}}\, dt \\
&= \int \frac{du}{C_B - \sigma^2 u} \\
&= -\frac{1}{\sigma^2} \log\bigl(C_B - \sigma^2 u\bigr) + C_A
\end{aligned}$$

This actually simplifies the A equation (37), to

$$\begin{aligned}
A &= C_A \exp\left(\frac{-1}{2} \log\bigl(C_B - \sigma^2 u\bigr) \right) \\
&= \frac{C_A}{\sqrt{C_B - \sigma^2 u}} \\
&= \frac{C_A}{\sqrt{C_B - \sigma^2 e^{2\gamma t}}} \ .
\end{aligned}$$

[1] There is a saying attributed to the brilliant theoretical computer scientist Donald Knuth: "Premature optimization is the root of all evil."

The actual expression for C_A depends on the actual expression for C_B, (34):

$$A(t) = \frac{C_A}{\sqrt{(2\gamma + \sigma^2)\, e^{2\gamma T} - \sigma^2 e^{2\gamma t}}} \ .$$

The final condition $A(T)$, finally, determines C_A:

$$\begin{aligned} 1 &= \frac{C_A}{\sqrt{2\gamma e^{2\gamma T}}} \\ \sqrt{2\gamma e^{2\gamma T}} &= C_A \\ A(t) &= \sqrt{\frac{2\gamma e^{2\gamma T}}{(2\gamma + \sigma^2)\, e^{2\gamma T} - \sigma^2 e^{2\gamma t}}} \\ A(t) &= \sqrt{\frac{2\gamma}{(2\gamma + \sigma^2) + \sigma^2 e^{-2\gamma(T-t)}}} \ . \end{aligned} \tag{38}$$

You can impress people with the final formula written out completely:

$$f(x,t) \ = \ \sqrt{\frac{2\gamma}{(2\gamma + \sigma^2) + \sigma^2 e^{-2\gamma(T-t)}}} \ e^{\frac{-\gamma x^2}{(2\gamma+\sigma^2)e^{2\gamma(T-t)} - \sigma^2}} \ .$$

Or, you could learn something useful by writing the approximate $A(0)$ for large T:

$$A(0) \approx \sqrt{\frac{2\gamma}{2\gamma + \sigma^2}} \ .$$

We can go further. There is an exponentially large range of x where the exponent is exponentially small (close to zero), which means that $f \approx A$ there.

This means that there is an exponentially large range of x values for which

$$f(x,0) = \mathrm{E}_{x,0}\Big[\, e^{-X_T^2/2} \Big] \approx \sqrt{\frac{2\gamma}{2\gamma + \sigma^2}} \ . \tag{39}$$

We conclude: Once T is large enough, and unless X_0 is too large, the expected value of $V(X_T)$ has a value that is almost independent of X_0 and T. All this algebra led to a simple qualitative conclusion that we really should try to understand ... next week.