Lecture notes by:
Matthias Kredler
Universidad Carlos III de Madrid
This version: July 2018

Tricks for exit times and exit probabilities

For one-dimensional diffusion processes, there are some nice tricks, and even closed-form solutions, for exit probabilities and expected exit times from intervals (see Section 1 of these notes), but also for densities (Section 3). These methods are especially useful when we have to determine if a diffusion whose volatility converges to zero at a boundary will ever hit that boundary (see Section 2). As far as I know, no equivalent results are available for multi-dimensional processes; however, some PDEs still hold, which can be exploited computationally (see Section 4). The notes mainly build on ideas laid out in the paper by Helland (1996), but try to bring them to a level that is accessible to mere mortals. The closed-form solutions for densities are taken from Luttmer's (2007) beautiful paper on firm dynamics.

1 Diffusions on an interval

1.1 General idea

The general idea that we will use throughout these notes is the one depicted in Figure 1: We will warp space (the vertical axis) and time (the horizontal axis) to make the the diffusion X_t in the left panel of Figure 1 look like the standard Wiener Process plotted on the right-hand side. We will then exploit that we know lots of things about the standard Wiener Process.

First, note that the process X_t on the left-hand side of the figure has an upward drift; it turns out that we can get rid of this drift by compressing the vertical axis. We have to shorten distances for high values of X_t but increase distances for low values of X_t. This warping of space is indicated by the vertical grid lines in the two graphs, which correspond to each other.[1]

Second, we see that the diffusion X_t on the left has higher volatility (it jiggles more) when X_t takes high values. It turns out that we can standardize volatility by stretching out the vertical axis (time) whenever the process X_t takes high values and by compressing it whenever X_t takes low values. This warping of time is indicated by the vertical grid lines, which again correspond to each other one-for-one in the left and the right panel.

Indeed, when you look closely, you see that the process W_t on the right is just a transformation of the process X_t from the left. You can check this by following X_t and verifying that indeed X_t crosses a grid line in the left graph

[1] The lowest vertical line on the left panel corresponds to the lowest vertical line on the right panel, the second-lowest to the second-lowest, and so forth.

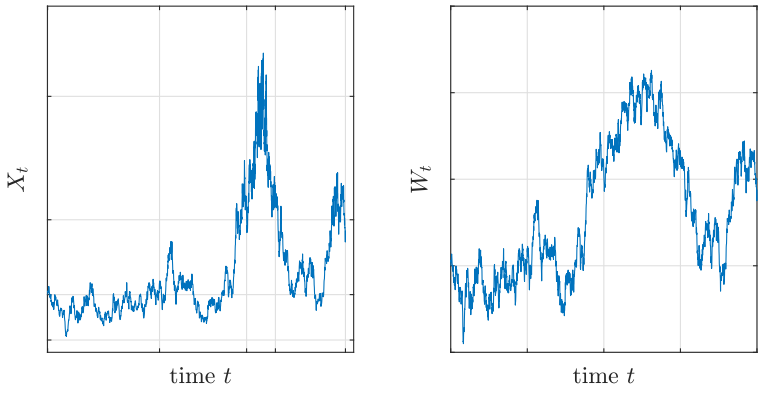


Figure 1: Realizations of a diffusion X_t and a corresponding standard Wiener Process W_t

exactly in the moment at which W_t crosses the corresponding grid line on the right side. Let's see now how we can formalize these simple geometric ideas mathematically.

1.2 Brownian Motion on the unit interval: Exit probabilities

We will start with the easiest-possible case: a random walk with constant variance on the unit interval. Let's define the diffusion process

$$\mathrm{d}X_t = \sigma \mathrm{d}W_t, \qquad \text{given } X_0 \in (a,b) = (0,1), \tag{1}$$

where W_t is the standard Wiener Process and where $\sigma > 0$ is the volatility of the process X_t. We can easily see that this stochastic differential equation (SDE) has the explicit solution $X_t = X_0 + \sigma W_t$. For $\sigma = 1$, this is just the standard Wiener Process.

To make things easier to grasp, let's think of X_t as the productivity of a firm. Let's suppose the following: If X_t ends up at the upper end of the interval, b, then the firm goes public (i.e. it is sold on the stock market) and the process ends there; this is the good outcome, in which the owners make a lot of money. The bad outcome is if the process X_t hits the lower bound of the interval, a; when this occurs, we assume that the firm shuts down and the owners get nothing – the bad outcome.

One interesting object is now the probability that the firm goes public given that we start the process at some $X_0 \in (a,b)$. Let $P(X_0)$ denote the probability that the process exits the interval (a,b) at b given that we start the process at X_0. The first thing to note is that X_t will exit the interval $[0,1]$ almost surely as $t \to \infty$; after all, the random variable X_t is normally distributed with

mean X_0 and variance $\sigma^2 t$, so the distribution fans out more and more and the probability that the process stays within the interval vanishes as t grows large. Thus we can infer that the firm goes broke eventually (i.e. that X_t exits at a) is $1 - P(X_0)$.

We will now try to get our hands on the function $P(\cdot)$ by exploiting the Markov property of the process X_t and state what is known as the *Kolmogorov Backward Equation* for the function $P(\cdot)$. This works as follows. Note that for any interior point $X_t \in (a, b)$, the process has zero probability to exit (a, b) over a short time interval $\mathrm{d}t$ (at least when we let $\mathrm{d}t$ be small enough). But since X_t is a Markov process, the probability $P(X_{t+\mathrm{d}t})$ only depends on the value that X takes at the end of the interval, and we can write:

$$P(X_t) = \mathbb{E}_t[P(X_{t+\mathrm{d}t})] \simeq \mathbb{E}_t\big[P(X_t) + P'(X_t)\mathrm{d}W_t + \tfrac{\sigma^2}{2}P''(X_t)\mathrm{d}t\big],$$

where the approximation in the last step follows from the Ito Formula. Now, since X_t has no drift, the stochastic term goes away, i.e. $\mathbb{E}_x[\mathrm{d}W_t] = 0$. We thus obtain the following restriction on the function $P(\cdot)$:

$$P''(x) = 0 \qquad \text{for all } x \in (a, b). \tag{2}$$

This is a second-order *ordinary differential equation* (ODE) for the unknown function $P(\cdot)$. It simply says that the function P cannot have any curvature and thus must be a straight line.[2] Also, the following boundary conditions must – quite obviously – be fulfilled:[3]

$$\lim_{x \to 0} P(x) = 0, \qquad\qquad \lim_{x \to 1} P(x) = 1.$$

The only function that is linear and fulfills these conditions is obviously the identity function, so the solution to our problem is

$$P(x) = x. \tag{3}$$

This solution is incredibly simple, it does not even depend on σ! So we will do whatever it takes to transform other diffusion processes into the form of a simple random walk in order to exploit this simplicity.

But before we start to look at other processes and other variables of interest (such as expected exit times), let us first build some more intuition for the simple case and let's consider how we can approximate the process X_t on a discrete grid.

[2] In higher dimensions, the condition $P'' = 0$ will be replaced by the Laplacian of the function $P : \mathbb{R}^n \to \mathbb{R}$ being zero. This means that the function P is *harmonic*, i.e. each function value is a local average over the values the function takes in its surroundings. Such functions are important in physics, for example in the study of heat transfer.

[3] In the discrete approximation in the next subsection, you will clearly see where the boundary condition comes from...

1.3 Approximation on a grid

We will now derive the function $P(x) = x$ for the exit probability at the top by approximating the random walk from Eq. (1) on a grid. To do this, we first cut the unit interval into n intervals of size $\Delta x_n = 1/n$. The $(n+1)$ grid points are denoted by $\{\tilde{x}_i\}_{i=0}^n \equiv \{i/n\}_{i=0}^n$. In general, it will hold true that when we let $n \to \infty$ we obtain in the limit the results that we obtained for the continuous-time diffusion. This works since the discrete stochastic process that we will construct converges to the true process in its stochastic properties.[4]

We let our approximating process $\tilde{X}_t$ start at the grid point$\tilde{X}_0$ that is closest to the original X_0. From there on, we let the process either go one grid point up or down, that is we set

$$\tilde{X}_{t+\Delta t_n} = \begin{cases} \tilde{X}_t + \Delta x_n & \text{with probability } p = \frac{1}{2}, \\ \tilde{X}_t - \Delta x_n & \text{with probability } p = \frac{1}{2}. \end{cases}$$

Now the question is: How should we choose the time increment, Δt_n? The increments of the true (continuous) process have variance σ^2 per unit of time, i.e. we have $\mathbb{E}_t[X_{t+s} - X_t]^2 = \sigma^2 s$. Thus, a reasonable thing to do is to set Δt_n such that

$$\mathbb{E}_t[\tilde{X}_{t+\Delta t_n} - \tilde{X}_t]^2 = (\Delta x_n)^2 = \sigma^2 \Delta t_n.$$

From this restriction, we obtain

$$\Delta t_n = \frac{\Delta x_n}{\sigma^2} = \frac{1}{(n\sigma)^2}.$$

This says that the more volatile the process is, the shorter we have to chose the time increments. Also, the finer our grid is, the shorter we have to make the time increments.

Now, let us derive the relationship that the function value $P(\tilde{x}_i)$ must have to the function values at the neighboring grid points, $P(\tilde{x}_{i+1})$ and $P(\tilde{x}_{i-1})$. We know that out of the paths that go up to $\tilde{x}_{i+1}$, a fraction $P(\tilde{x}_{i+1})$ will exit the interval at 1 eventually. Similarly, out of the paths that go down one step, a fraction $P(\tilde{x}_{i-1})$ will end up at 1 eventually. We thus have

$$P(\tilde{x}_i) = \tfrac{1}{2}P(\tilde{x}_{i+1}) + \tfrac{1}{2}P(\tilde{x}_{i-1}) = \mathbb{E}_t\big[P(\tilde{x}_{t+1})\big] \qquad \text{for } i = 1, \ldots, n-1.$$

So $P(\tilde{x}_i)$ is a local average of the function values P takes in its immediate vicinity. Seen in another way, this equation says that the process $P(\tilde{X}_t)$ has the martingale property (just as its continuous-time counterpart $P(X_t)$ does): The expectation of future values $P(\tilde{x}_{t+j})$ is identical to $P(\tilde{x}_t)$ itself. Note, by the way, that the above equation also holds for the two grid points adjacent to the boundaries, $\tilde{x}_1$ and $\tilde{x}_{n-1}$, since we have $P(\tilde{x}_0) = P(0) = 0$ and $P(\tilde{x}_n) = P(1) = 1$.

[4] This, and nothing more, is the intuition behind the Markov-chain approximation method that is used to solve PDE and SDE problems numerically.

Another way of looking at the above equation is that the gain in probability of going public when the firm goes up one grid point has to be offset by the loss in the probability when the process $\tilde{x}_t$ goes down one grid point. We can re-write the last equation as

$$\big[P(\tilde{x}^+) - P(\tilde{x})\big] - \big[P(\tilde{x}) - P(\tilde{x}^-)\big] = 0,$$

where we denote the grid point just above $\tilde{x}$ by $\tilde{x}^+$ and the grid point just below by $\tilde{x}^-$. If we divide the above equation by Δx_n, we obtain

$$\frac{P(\tilde{x}^+) - P(\tilde{x})}{\Delta x_n} = \frac{P(\tilde{x}) - P(\tilde{x}^-)}{\Delta x_n}.$$

Note that $[P(\tilde{x}^+) - P(\tilde{x})]/\Delta x_n$ is the *right difference quotient* – in numerical methods, this is what we use as an approximation for the first derivative of $P(\cdot)$ when we approximate a function $P(\cdot)$ on a discrete grid. Thus, the last equation says that the slope of P, ΔP, should not change throughout the interval. We can, indeed, divide the last equation again by Δx_n to obtain

$$\frac{\frac{P(\tilde{x}^+)-P(\tilde{x})}{\Delta x_n} - \frac{P(\tilde{x})-P(\tilde{x}^-)}{\Delta x_n}}{\Delta x_n} = \underbrace{\frac{P(\tilde{x}^+) - 2P(\tilde{x}) + P(\tilde{x}^-)}{(\Delta x_n)^2}}_{\to P''(\tilde{x}) \text{ as } n\to\infty} = 0.$$

This equation says that the *second difference quotient* at $\tilde{x}_i$ (which is the discrete approximation for the second derivative) should equal zero at all grid points. We recognize that this last equation is nothing but the discrete translation of the second-order ODE $P''(x) = 0$ from Eq. (2). Obviously, the only solution is again the identity function $P(\tilde{x}_i) = \tilde{x}_i$. We also see clearly now that at the boundaries, it must be that $P(\tilde{x}_1)$ and $P(\tilde{x}_{n-1})$ converge to zero and one, respectively, as $n \to \infty$. This must be the case since $P(\tilde{x}_{n-1})$ is an average of 1 and $P(\tilde{x}_{n-2})$, and $P(\tilde{x}_1)$ is an average of 0 and $P(\tilde{x}_2)$.

In the discrete-grid approximation, it is also easier to see why the variance σ does not matter for the exit probability $P(x)$: The approximating process $\tilde{X}_t$ takes the same form any volatility; we only make the steps take a longer time when σ is low. But obviously, it does not matter how fast uncertainty is resolved if we are only interested in the probability that the process exits at the top eventually.

1.4 Taking expectations of any function: The Green Function as the expected-local-time measure

Kolmogorov Backward Equation for general test functions. Volatility does not matter for the success probability of the firm in our example, but surely volatility matters for *how fast* the uncertainty about the success of the firm is resolved. It will also matter for other functions of interest, such as expected profits. Let us assume that flow profits of the firm at time t are given by some

function $\pi(X_t)$. How would we go about computing expected profits of our firm before it exits the interval?

Just as before for the case of $P(\cdot)$, the answer is the Kolmogorov Backward Equation. However, we will be more ambitious than just getting the expectation of the function $\pi(\cdots)$: Let us set as a goal to calculate the expectation over *any* function $f(X_t)$ that is defined as a function of firm productivity X_t; we call these functions *test functions*. In our firm example, this function could describe flow profits, but it could also be the number of employees the firm has, or investment that depends on productivity X_t, or any moment of these (say profits squared) etc.

Let us first define the random variable T as the random time at which X_t first exits the unit interval.[5] We then define the *expected exit time* as the function $S(X_0) \equiv \mathbb{E}_0[T] = \mathbb{E}[T|X_0]$. Finally, for a general function $f : (0,1) \to \mathbb{R}$, we define the expectation of f given the starting value X_0 as the function

$$F(X_0) = \mathbb{E}_0\Big[\int_0^T f(X_t)\mathrm{d}t\Big].$$

Note here that for the specific test function $f(x) = 1$, this gives us the expected exit time $S(X_0)$, since $\mathbb{E}_0[\int_0^T 1\mathrm{d}t] = \mathbb{E}_0[T]$. You may ask why we do not discount expected profits, say at a rate ρ. We will leave this extension as an exercise; it turns out that omitting the discounting will make it possible to derive a measure of the expected time that the process spends at each point in the interval, so let's do this first.

How will it be possible to get our hands on the expected amount of time that X_t spends at different points in the state space? Note that in the last equation, we should read the expectation operator as an integral over all paths $\{X_t\}_{t=0}^T$ that the process can take, weighted by the probabilities that these paths occur. We will now try to change from integration over *paths* to integration over *states* $x \in [0,1]$. To be precise, we ask if it possible to write $F(\cdot)$ as an integral of the form

$$F(X_0) = \int_0^1 f(x)G(X_0, x)\mathrm{d}x$$

for some function $G(X_0, x)$. Here, $G(X_0, x)$ would be the expected amount of time that the process will spend at x given that we start the process at X_0. $G(X_0, \cdot)$ is a density over the state space that we call *expected local time* at x.[6]

It turns out that in the one-dimensional case, it is possible to find the function $G(X_0, \cdot)$ in closed form in many examples using the wizardry of *Green Functions*. Let's see now how this works. As before, we start by stating the Kolmogorov Backward Equation to find out more about the properties of the function $F(\cdot)$. Over a short time horizon Δt, we have, for any interior $X_0 \in (a, b)$,

[5] Since the Wiener Process is a very weird object and jiggles wildly even at very short horizons, formally we have to define this random variable, a *stopping time*, as $T \equiv \inf\{t : X_t \geq 1 \text{ or } X_t \leq 0\}$; the simpler $\min\{t : X_t \geq 1 \text{ or } X_t \leq 0\}$ is not a well-defined object.

[6] If we discounted profits, this would be *discounted expected local time* at x.

that

$$F(X_t) = f(X_t)\mathrm{d}t + \mathbb{E}_t\big[F(X_{t+\mathrm{d}t})\big] \simeq f(X_t)\mathrm{d}t + F(X_0) + \frac{\sigma^2}{2}F''(X_0)\mathrm{d}t,$$

where we have used the Ito Formula in the second step. Again, the term in $F'(X_t)\mathrm{d}X_t$ has dropped out since X_t has no drift. Since X_0 was arbitrary, we can re-arrange the above equation to obtain the following 2nd-order ODE for $F(\cdot)$:

$$F''(x) = -\frac{2}{\sigma^2}f(x) \qquad \text{for all } x \in (0,1). \tag{4}$$

The boundary conditions are again $F(0) = F(1) = 0$; no profits can be made any more once the process has exited the interval. At first, Equation 4 looks impossible to solve – after all, $f(\cdot)$ may be *any* function! But let's not give up so quickly...

Discretizing. Let us now pause and consider our discretized process to get some idea on how to solve the ODE (4). In the discrete world, total expected profits given initial state $\tilde{X}_0$ are given by the sum

$$F(\tilde{X}_0) = \mathbb{E}_0 \sum_{i=1}^{N} f(\tilde{X}_i)\Delta t_n,$$

where N is the (random) number of steps that the process takes until arriving at either 0 or 1 and where $\tilde{X}_i$ is the value that $\tilde{X}$ takes at step i. Again, $\Delta t_n = (\Delta x_n)^2/\sigma^2$ is the time increment, which depends on the process' volatility and the mesh size of the grid. We have to multiply flow profits $f(\tilde{X}_i)$ by the length of time that the firm enjoys these profits, that's why this term shows up in the sum. We note that the above sum is over all *paths* of the process $\{\tilde{x}_i\}_{i=1}^N$, just as our initial integral was over *paths* of the process X_t.

As in the continuous case, we can use the Markov property to write total profits recursively as a sum of the profits in the current time period plus expected profits from the next period on. Since $\tilde{X}_t$ is a martingale, we have

$$F(\tilde{x}_i) = f(\tilde{x}_i)\Delta t_n + \tfrac{1}{2}F(\tilde{x}_{i+1}) + \tfrac{1}{2}F(\tilde{x}_{i-1}) \qquad \text{for } i = 1, 2, \ldots, n-1.$$

Since we have one equation for all interior grid points, this is a system of $n-1$ linear equations in the $n-1$ unknowns $\{F(\tilde{x}_i)\}_{i=1}^{n-1}$. In matrix form, the system can be written as

$$\underbrace{\frac{1}{(\Delta x_n)^2}\begin{bmatrix} -2 & 1 & 0 & \ldots & & \\ 1 & -2 & 1 & 0 & \ldots & \\ 0 & 1 & -2 & 1 & 0 & \ldots \\ & & \vdots & & & \\ 0 & \ldots & 0 & 1 & -2 & 1 \\ & & \ldots & 0 & 1 & -2 \end{bmatrix}}_{\equiv \mathbf{H}} \underbrace{\begin{bmatrix} F(\tilde{x}_1) \\ F(\tilde{x}_2) \\ \\ \vdots \\ \\ F(\tilde{x}_{n-1}) \end{bmatrix}}_{\equiv \mathbf{F}} = \underbrace{-\frac{2}{\sigma^2}\begin{bmatrix} f(\tilde{x}_1) \\ f(\tilde{x}_2) \\ \\ \vdots \\ \\ f(\tilde{x}_{n-1}) \end{bmatrix}}_{\equiv \hat{\mathbf{f}}}, \tag{5}$$

where we have used our formula $\Delta t_n = \Delta x_n/\sigma^2$ for the time increment; we can write this system compactly as $\mathbf{HF} = -\hat{\mathbf{f}}$. Now, stare at the $(n-1)$-square matrix $\mathbf{H}$. You can recognize that $\mathbf{H}$ is the second-difference operator on vectors of length $(n-1)$![7] So the system (5) is obviously the discrete counterpart to the second-order ODE we found in (4) for the continuous process.

Our task now is to find the inverse of the matrix $\mathbf{H}$ to solve for the values $\{F(\tilde{x}_i)\}_{i=1}^{n-1}$. We won't do this by brute force but pause and think for a moment. In principle, we are interested in finding the solution $\mathbf{F}$ to the system (5) given *any* vector $\hat{\mathbf{f}}$. What we will try now is to solve the system first for setting $\hat{\mathbf{f}}$ equal to the $n-1$ unit vectors in $\mathbb{R}^{n-1}$ and then to combine these solutions for vectors $\hat{\mathbf{f}}$ of arbitrary nature. Let $\mathbf{e}_i = [0, \ldots, 1, \ldots, 0] \in \mathbb{R}^{n-1}$ be the ith unit vector of length $n-1$ and denote by $\mathbf{E}_i$ the solution of (5) for $\hat{\mathbf{f}} = \mathbf{e}_i$. It turns out that if we can do this, we are done... We can then just combine the solutions linearly to construct any solution! To see this, note that we can write $\hat{\mathbf{f}}$ as a combination of the unit vectors, i.e. as

$$\hat{\mathbf{f}} = \hat{f}_1\mathbf{e}_1 + \cdots + \hat{f}_{n-1}\mathbf{e}_{n-1},$$

where $\hat{f}_i$ denotes the ith component of vector $\hat{\mathbf{f}}$. Then it must be that the vector

$$\mathbf{F} = \hat{f}_1\mathbf{E}_1 + \cdots + \hat{f}_{n-1}\mathbf{E}_{n-1}$$

solves the system (5) – just try it out to see that this works. Indeed, the vectors $\mathbf{E}_i$ are nothing but the columns of the matrix $\mathbf{H}^{-1}$, the inverse of $\mathbf{H}$ – so in fact we are just re-doing some fundamental linear algebra here!

But what do the fundamental solutions (the vectors $\mathbf{E}_i$ that solve $\mathbf{HE}_i = \mathbf{e}_i$) look like? Since $\mathbf{H}$ is the second-difference operator, when plotting the components of the vector against the position $i = 1, \ldots, n-1$, the vector $\mathbf{E}_i$ should have a spike at position i, but at all other positions the vector should be "flat" (i.e. the slope shouldn't change). Specifically, the spike at i should be such that the change in the slope equals -1. Also, the vectors E_i should start at zero and end at zero again. Doing the maths, it turns out that setting

$$E_i(\tilde{x}) = \begin{cases} \tilde{x}(1-\tilde{x}_i) & \text{if } \tilde{x} \leq \tilde{x}_i, \\ \tilde{x}_i(1-\tilde{x}) & \text{if } \tilde{x} > \tilde{x}_i \end{cases}$$

does the job. Figure 2 shows what these vectors (drawn as functions of x here) look like for a grid with $n = 10$.

We now combine the solutions for the $n-1$ basis vectors to construct our solution for a general vector $\hat{\mathbf{f}}$ as

$$F(\tilde{x}_i) = \sum_{j=1}^{n-1} E_i(\tilde{x}_j)\hat{f}(\tilde{x}_j) = \sum_{j=1}^{n-1} \underbrace{\frac{2E_i(\tilde{x}_j)}{\sigma^2}}_{\text{expected local time at } \tilde{x}_j \text{ given } \tilde{x}_i} f(\tilde{x}_j). \qquad (6)$$

[7] Note that $\mathbf{H}$ also gives the second difference for the first and last element of the vector since we have $F(\tilde{x}_0) = F(\tilde{x}_n) = 0$.

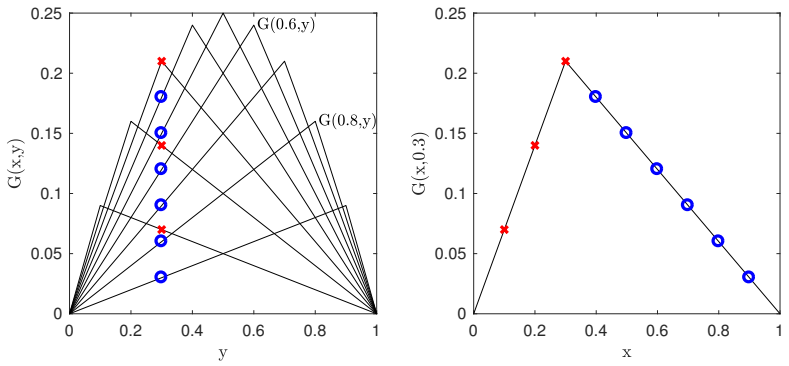


Figure 2: Green's function on unit interval and weights for $y = 0.3$

It is claimed that in the last expression, the term $2E_i(\tilde{x}_j)/\sigma^2$ is the expected amount of time the process $\tilde{X}_t$ will spend at grid point $\tilde{x}_j$ given that the process starts at $\tilde{X}_0 = \tilde{x}_i$. Why is this? Just set the test function to be the jth unit vector, i.e. set $\mathbf{f} = \mathbf{e}_j$. Then, it clearly follows that $2E_i(\tilde{x}_j)/\sigma^2$ is expected time spent at grid point j. Every time the process visits $\tilde{x}_i$, we count the time spent there, but we don't count up time when visiting any other grid point. As is intuitive, a high σ decreases expected local time at all grid points: The higher volatility, the faster the process will leave the interval, and thus expected local time is shorter for all points of the grid.

It turns out that for the discretized process, we can also say something about *how often* the process will visit a specific grid point. When we divide expected local time at $\tilde{x}_j$ by the length of the time interval, $\Delta t_n = 1/(n\sigma)^2$, we obtain the expected number of times $N(\tilde{x}_i, \tilde{x}_j)$ that the process visits $\tilde{x}_j$ if started at $\tilde{x}_i$ as

$$N(\tilde{x}_i, \tilde{x}_j) = 2E_i(\tilde{x}_j)n^2. \tag{7}$$

Figure 3 shows a histogram for $N(0.3, \cdot)$. We see that the histogram inherits the spike shape of the Green Function from the last graph. Crucially, our formula (7) for $N(\cdot)$ does not depend on the variance of the process, σ^2. Since we use the same binomial process for any value of σ (we just make the time increments shorter), this actually *must* be the case, once you start thinking about it. We will use exactly this insight again later when we start to warp time for processes with time-varying volatility...

Back to the continuous case: *Green's Function.* Can we transfer our solution method from the discrete to the continuous world? It turns out that the answer is "yes", and that we can discover some breathtaking mathematical beauty when doing so. Let us go back again to the ODE (4) that we wanted to solve, but re-write it slightly:

$$\frac{\partial^2}{\partial x^2} F(x) = -\hat{f}(x), \tag{8}$$

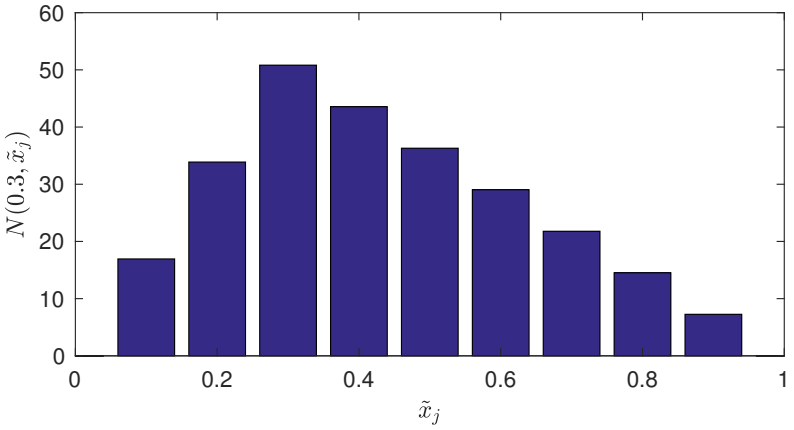


Figure 3: Expected number of visits for discrete process given $\tilde{X} = 0.3$, $n = 10$.

where we define $\hat{f}(x) \equiv 2f(x)/\sigma^2$. We note that the operator $\mathcal{H} \equiv \frac{\partial^2}{\partial x^2}$ plays the role that $\mathbf{H}$ played in the discrete case in the system (5). Indeed,the mapping $\mathcal{H}$ (which maps functions to functions) is a linear mapping, just as the matrix $\mathbf{H}$ is: If we apply $\mathcal{H}$ to linear combinations of two functions, the outcome will always be a linear combination of the outcomes we would have obtained when applying the mapping separately to the two functions, i.e. we have $\partial^2/\partial x^2[g(x)+h(x)] = g''(x) + h''(x)$.

We now would like to proceed the same way as we did in the discrete case with the system of equations (5), i.e. we would like to solve the system when setting $\hat{f}(\cdot)$ to the unit vectors. But this requires us to take a stand what the unit vectors are in the space of all functions defined on the unit interval. It turns out that these unit vectors are the famous *Dirac deltas*, $\delta_x(\cdot)$. There is one such delta function for each point $x \in (0, 1)$. The Dirac delta function $\delta_x(\cdot)$ is like a probability distribution that has a unit mass points at point x, but zero

mass everywhere else.[8] Each Dirac delta corresponds to the unit vector $\tilde{x}_i$ for some position $\tilde{x}_i$ on the grid, but now there are infinitely many of them.

Now, let's go back to our ODE. Setting $\hat{f}(x) = \delta(x)$ in Eq. (8) means the following: The second derivative of F, $\mathcal{H}F$, should be zero for all $\tilde{x} \neq x$, but F should have a sharp spike at x, In fact, its derivative should change by -1 at x, i.e. we should have $F'_+ = F'_-(x) - 1$, where $F'_+(x)$ and $F'_-(x)$ denote the right and left derivative of F at x. Finally, the boundary conditions $F(0) = F(1) = 0$ should be fulfilled.

It is not too hard to construct the functions that fulfill these criteria. Indeed, the solution to (8) for all $\{\delta(x)\}_{x\in(0,1)}$, is given by *Green's Function*, which looks exactly like our vectors $\{E_i\}$ that are plotted in Figure 2. So we define *Green's Function* as[9]

$$G(x,y) = \begin{cases} x(1-y) & \text{if } x \leq y, \\ (1-x)y & \text{if } x > y. \end{cases}$$

[8]For the mathematically inclined: Dirac deltas are not functions in the usual sense; we may call them *distributions* or *generalized functions*. Dirac deltas are elements of the *dual space* of C^0, where C^0 denotes the space of continuous functions on the unit interval. A dual space is the space of linear functionals that we can define on elements of another space, in our case C^0. Note that we can define the linear functional $\delta_x(\cdot)$ as follows: The functional δ_x just throws back the value that a function $f \in C^0$ (the input into the functional) takes at a some (fixed) point $x \in [0,1]$. Formally, let us define the Dirac delta δ_x by the following equation:

$$\langle \delta_x | f \rangle = \int_0^1 \delta_x(y) f(y) \mathrm{d}y = f(x).$$

Here, the notation $\langle \delta_x | f \rangle$ means that we apply the linear functional δ_x to a function f. The second step defines a kind of generalized integral that allows for mass points. So now, why is δ_x a linear functional? When we apply it to linear combinations $f + \lambda g$ of continuous functions f, g (λ being a scalar), we obtain exactly the same results as when applying the functional separately (check this):

$$\langle \delta_x | f + \lambda g \rangle = f(x) + \lambda g(x) = \langle \delta_x | f \rangle + \lambda \langle \delta_x | g \rangle.$$

Of course, also integration against other functions in C^0 gives us a (maybe conceptually simpler) linear functional: $\langle g | f \rangle = \int_0^1 g(y) f(y) \mathrm{d}y$. But the Dirac deltas show us that the dual of C^0 is way larger than the space of continuous functions itself. In fact, any probability distribution defined on the unit interval defines one functional of this kind. Any probability distributions can be described by a cumulative distribution function (cdf), i.e. an increasing function that satisfies $F(0) \geq 0$ and $F(1) = 1$. The Dirac deltas δ_x are characterized by cdfs that jump from zero to 1 at x (these are called *heavy-side step functions*); so you can think of δ_x as the derivative of the function that jumps from zero to one at x , or as the probability density function (pdf) of a probability that puts all mass on one point. Of course, any differentiable cdf $F(\cdot)$ defines a linear functional on C^0; these functionals can be represented as integration against their pdf, $f(x) \equiv F'(x)$. However, also some weirder objects than differentiable cdfs and Dirac deltas are in the dual of C^0. Look up the *Cantor Function* (or *Devil's Staircase*) on Wikipedia, a classical example in Real Analysis, for a weird kind of cdf that has no mass points but does not allow a density representation either.

[9]There are different Green Functions for different ODE or partial-differential-equations (PDE) problems, each of which has to be constructed for the specific problem at hand. The formula given here is only valid for 2nd-order ODEs on the unit interval with boundary conditions $F(0) = F(1) = 0$.

Then the solution to our problem (think: expected profits given $x = X_0$) is

$$F(x) = \int_0^1 \underbrace{\frac{2}{\sigma^2} G(x,y)}_{\text{expected local time at } y\in(0,1) \text{ given } X_0=x} f(y)\mathrm{d}y \quad \text{for all } x \in (0,1).$$

You can check that this formula really gives us the solution to the ODE (4) – just try it out and take the derivative of $F(\cdot)$ with respect to x to see that the formula works! The comparison to the discrete case shows that the Green Function $G(\cdot,\cdot)$ is some kind of matrix for functions (also called a *kernel*): It is a linear mapping – it maps $f(\cdot)$ to $F(\cdot)$ – that is the inverse of the linear operator $\mathcal{H} = \partial^2/\partial x^2$! This should be somewhat breathtaking, at least for the mathematically inclined...

Expected local time. Again, as in the discrete case, it must be that the "basis solutions" $G(x,y)$ give us the expected amount of time that the process X_t will spend at y before exiting $[0,1]$, given that we start the process at $X_0 = x$. We can see this by setting test functions $f(\cdot)$ that are zero everywhere but take the value 1 in a small neighborhood of some fixed point y (in fact, we could also take Dirac deltas for f).

Again, notice the simplicity of the solution: Expected local time $G(X_0,\cdot)$ is a function that decreases linearly towards zero towards both end points of the interval and is inversely proportional to the variance of the process, σ^2. So again, we will do whatever it takes to press more complicated processes into this simple scheme...

Exercise 1: What if we want to calculate *discounted* integrals over functions, i.e.

$$F(X_0) = \mathbb{E}_0 \int_0^T e^{-\rho t} f(X_t)\mathrm{d}t?$$

Re-do the steps in the previous section and find the Green Function $G_\rho(x,y)$ for the discounted case (i.e. the function that gives us discounted expected local time at y given a starting value $X_0 = x$ for the diffusion). Hint: The exponential functions of type $e^{\alpha x}$ are a good candidate to solve ODEs of type $f''(x) = f(x) \times const.$

Exercise 2: Consider the space C^1 of continuously differentiable functions f on the unit interval that vanish at 0 (i.e. that fulfill $f(0) = 0$). Suppose that we are interested in solving the first-order ordinary differential equation (ODE)

$$\frac{\partial}{\partial x} R(x) = r(x),$$

where $r(\cdot)$ is a given function. Assume that $r(\cdot) \in C^0$, where C^0 is the space of continuous functions on the unit interval. Follow the same steps as in the last section to find the inverse to the linear operator $\mathcal{K} = \frac{\partial}{\partial x}$ (which maps C^1 to C^0, where C^1 denotes the space of continuously differentiable functions on the unit interval). The solution will be somewhat of a triviality, but this exercise helps you to grasp conceptually what we have been doing in this section. Hint: Look up the *heavy-side step functions* on Wikipedia if you don't make progress; also look them up if you can make progress.

1.5 Expected exit time from the unit interval

Given what we have done in the last section, it would now be easy to find the expected exit time $S(X_0)$ of the process X_t from the unit interval, $(a, b) = (0, 1)$: We just have to set $f(x) = 1$ and use Green's Function. However, it may be useful for you to derive the result by re-doing the steps above for the special case $f(x) = 1$ in order to recapitulate (otherwise, just skip this subsection and make sure you understand the equations)...

As always, over a short time horizon Δt, the process X_t almost surely does not exit the interval, thus by the Ito Formula we have

$$S(X_t) = \Delta t + \mathbb{E}_t[S(X_{t+\Delta t})] \simeq \Delta t + S(X_t) + \tfrac{\sigma^2}{2} S''(X_t)\Delta t.$$

Dividing by Δt and taking limits gives us the second-order ODE

$$S''(x) = -\frac{2}{\sigma^2},$$

which says that the function S must be concave with a constant curvature. The boundary conditions are $S(0) = S(1) = 0$. We see that the function

$$S(x) = \frac{x(1-x)}{\sigma^2} \tag{9}$$

fulfills these requirements and is thus the solution to our problem – you should verify that the same result obtains by using Green's Function. As is intuitive, a higher volatility decreases the expected exit time. Also, the expected exit time is symmetric around $x = \frac{1}{2}$, the middle of the interval, at which it is maximal.

1.6 Changing the interval: Affine transformations

The first generalization of our results concerns changes in the size of our interval: We now assume that the interval bounds are given by arbitrary numbers $a < b$. For the time being, we maintain the assumption that X_t follows the simple random walk with constant volatility given in the SDE (1). This turns out to be fairly simple: In a sense, we are just stretching the unit interval and shift it around in space, i.e. we apply an *affine* transformation to the process and the interval.[10]

Exit probabilities: Re-scaling the process. We will first find the probability $P_{ab}(x)$ that the process exits the interval at the upper bound b. By the same arguments as in Section 1.2, the function P_{ab} must satisfy $P''_{ab}(X) = 0$ and the boundary conditions $P_{ab}(a) = 0$ and $P_{ab}(b) = 1$. The only function that fulfills these conditions is a straight line that connects the two boundary points, i.e.

$$P_{ab}(X) = \frac{X - a}{b - a}. \tag{10}$$

[10] Affine transformations are functions of the form $f(x) = \alpha + \beta x$, where α, $\beta \in \mathbb{R}$.

We will now see that we would have obtained the same solution by re-scaling the process through an affine transformation. To see this, let us define another process, $\hat{X}_t$, by

$$\hat{X}_t = L(X_t) \equiv \frac{X_t - a}{b - a},$$

where $L(\cdot)$ is the affine transformation. Note that $\hat{X}_t$ lives on the unit interval. The transformed process has increments of size $\mathrm{d}\hat{X}_t = \sigma \mathrm{d}W_t/(b-a)$ and thus a different variance, namely

$$\hat{\sigma}^2 \equiv \frac{\sigma^2}{(b-a)^2}.$$

Applying the formula (3) that we had found for the exit probability at the top in the case of the unit interval to the scaled process $\hat{X}_t$, we find exactly the same result as before, namely

$$P(\hat{X}) = \hat{X} = \frac{X - a}{b - a} = P_{ab}(X).$$

Expected local time and a generalized Green Function. To find expected local time, the Kolmogorov Backward Equation for the expectation of an arbitrary function $f(\cdot)$ reads, as before,

$$F''(x) = -\frac{2}{\sigma^2} f(x)$$

and has boundary conditions $F(a) = F(b) = 0$. Again, we can solve this ODE using a Green Function, which for a general interval (a, b) takes the form

$$G_{ab}(x, y) = \begin{cases} \frac{(x-a)(b-y)}{b-a} & \text{if } x \leq y, \\ \frac{(y-a)(b-x)}{b-a} & \text{if } x > y. \end{cases} \tag{11}$$

Check that indeed this function has the spike shape that we require, i.e. that it (i) satisfies $G_{ab}(x, a) = G_{ab}(x, b) = 0$, (ii) its second derivative in y is zero for $y \neq x$ and that (iii) the change in the first derivative in y at $y = x$ is of size 1. Given this Green Function, the solution to the ODE for $F(\cdot)$ is

$$F(x) = \int_a^b \underbrace{\frac{2G_{ab}(x, y)}{\sigma^2}}_{\text{expected local time at } y} f(y)\mathrm{d}y.$$

Exercise: Derive the Green Function and expected local time in an alternative fashion: Apply our formulae for the unit interval to the transformed process $\hat{X}_t$. Hint: Don't forget to apply the formula for the change of variable when you switch from integration over $\tilde{x}$ to integration over x.

Expected exit time. Denote the expected exit time from the interval $[a, b]$ given initial condition X_0 by $S_{ab}(X_0)$. The Kolmogorov Backward Equation

gives us again the ODE $S''_{ab}(X) = -2\sigma^{-2}$ with boundary conditions $S_{ab}(a) = S_{ab}(b) = 0$. It turns out that the function

$$S_{ab}(x) = \frac{(x-a)(b-x)}{\sigma^2} \tag{12}$$

fulfills these conditions and looks very similar to the solution for the unit interval. Again, let's check that applying the formula (9) for the unit interval to the re-scaled process, $\hat{X}_t = (X_t - a)/(b-a)$, works. Indeed, we have

$$S(\hat{X}_0) = \frac{\hat{X}_0(1-\hat{X}_0)}{\hat{\sigma}^2} = S_{ab}(X_0),$$

which is the same as the formula in Eq. (12).

1.7 Changing volatility: Warping time

Extension to general martingale process. We now consider a more general martingale (i.e. a diffusion process without drift) on the unit interval $(a, b) = (0, 1)$. Our SDE is now

$$\mathrm{d}X_t = s(X_t)\mathrm{d}W_t,$$

where $s(x) > 0$ is the volatility of the process at x. In this section, we assume that the function $s(\cdot)$ is bounded away from zero, i.e. we require that $s(x) \geq \epsilon$ for all $x \in (0, 1)$ for some $\epsilon > 0$. This is sufficient to ensure that X_t always reaches both end points of the interval with positive probability.[11]

The upper left panel of Figure 4 plots two realizations of such a process for the volatility function $s(x) = \sigma x$ and $\sigma = 1$ with starting point $X_0 = 1$. Volatility increases linearly in the state X_t, which clearly manifests itself in the sample paths: They are jiggering less the closer they come to the horizontal axis.

Changing the time measure. The idea now is the following: We will compress the time axis whenever volatility is low and we expand it when volatility is high. Imagine this as defining a *subjective time scale* that a traveler experiences who moves on top of the process X_t. When volatility is high, innovations are large and many events are happening per time unit; things get more exciting and a given amount of time feels longer to our traveler. When shocks are smaller, however, fewer events are happening and subjectively the traveler feels that time is passing faster (let's imagine that boredom is not a problem for our traveler).

To make this idea mathematically precise, let us define a new (subjective) time scale $\tau(t)$ as a funtion of the original time scale, t. $\tau(t)$ is an increasing function of time; also subjective time can only run forward. But the slope of the function $\tau(\cdot)$ will vary with how fast uncertainty is resolved. Formally, we

[11] In Section 2, we will study the interesting case in which $\lim_{x\to 0} s(x) = 0$; then, it is not so clear if the process X_t can reach the bounds of the interval, as we will see.

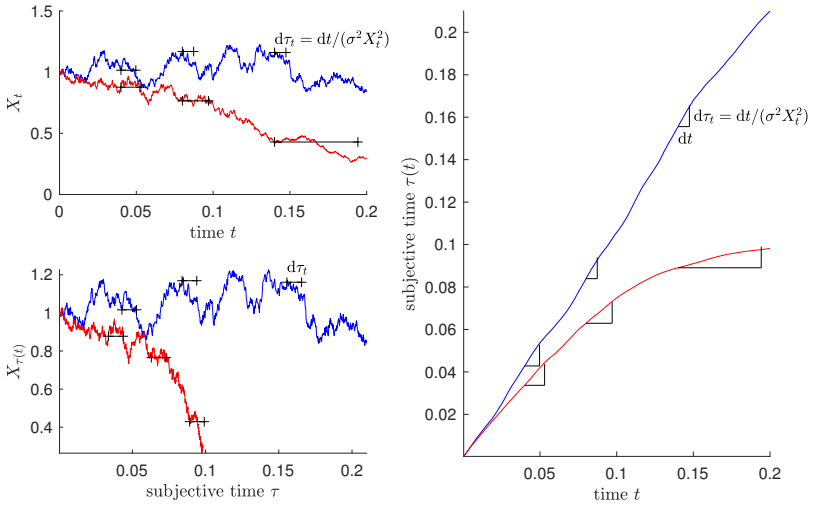


Figure 4: Warping time for the process $\mathrm{d}X_t = \sigma X_t \mathrm{d}W_t$ with $\sigma = 1$

define the function $\tau(t)$ such that the same amount of uncertainty is resolved in a fixed amount of subjective time, $\mathrm{d}\tau$. Specifically, we require that

$$\begin{aligned}\mathbb{E}_t\big[X_{\tau(t+\mathrm{d}t)}) - X_{\tau(t)}\big]^2 &= \mathbb{E}_t\big[s(X_t)(W_{\tau(t+\mathrm{d}t)} - W_{\tau(t)})\big]^2 \\ &= s(X_t)^2\big[\underbrace{\tau(t+\mathrm{d}t) - \tau(t)}_{=\mathrm{d}\tau}\big] \qquad \stackrel{!}{=} 1\mathrm{d}t,\end{aligned}$$

where the first step just uses the SDE for the process X_t, the second step makes use of the fact that the increments of the standard Wiener process have unit variance, i.e. $\mathbb{E}_t[W_{t+\mathrm{d}t} - W_t]^2 = \mathrm{d}t$, and the the last equality imposes the normalization that the same amount of uncertainty is resolved for any subjective time increment, $\mathrm{d}\tau$. Re-arranging the above, we find that the function $\tau(\cdot)$ must have slope

$$\tau'(t) = \left.\frac{\mathrm{d}\tau}{\mathrm{d}t}\right|_t = \frac{1}{s(X_t)^2} \equiv m(X_t),$$

where the function $m(\cdot)$ gives us the change in the time measure. The right panel of Figure 4 shows how subjective time τ evolves as a function of (real) time t for the two sample paths. Note that the slope $m(X_t)$ of the two lines in the right panel tells us how short intervals of real time are transformed into subjective time.[12]

[12] If you ever wondered what a function looks like that has continuous first derivative but whose second derivative does not exist, the right panel of Figure 4 may be the best chance you get to see one. Note that the paths in the left-upper panel of the figure are, by definition, the derivatives of the functions plotted on the right-hand side; these derivatives are continuous but nowhere differentiable, since they inherit the properties of the Wiener Process.

Finally, the lower-left panel of Figure 4 shows what the two sample paths look like when we plot them against their subjective time scales. The black brackets in the upper-left and lower-left panels mark intervals that are of the same "subjective length" $\mathrm{d}\tau$. Since the lower (red) sample path has lower volatility, it is compressed and becomes shorter than the upper (blue) path. We also see that the transformation into subjective time makes the processes look like a standard Wiener process with constant volatility. Interestingly, the red (lower) path seems to be grinding to a halt; we will later get back to this and discover that it can indeed occur that subjective time stays finite while real time goes to infinity – a phenomenon that can arise for processes whose volatility approaches zero at a boundary.

As before, we will now set out to calculate expectations of functions $f(X_t)$ and obtain solutions for objects such as expected profits and expected exit times. We expect the process X_t to be trapped for some time in areas with low volatility, leading to longer exit times and a higher weighting of such areas when it comes to expected profits, whereas the expected exit time should be shorter and the weighting for profits should be lower in areas with high volatility, where uncertainty is resolved earlier. Formally, we can use again the Ito Formula on our function $F(X_0) = \mathbb{E}_0 \int_0^t f(X_t)\mathrm{d}t$ to find the second-order ODE

$$F''(x) = -\frac{2}{s(x)^2} f(x). \tag{13}$$

The boundary conditions for the ODE are again $F(0) = F(1) = 0$. When setting $f(x) = 1$, then this is the ODE that the expected exit time $S(x)$ has to solve. For this case, Eq. (13) says that the higher the volatility $s(x)$ is at x, the faster uncertainty is resolved, and the flatter (or less concave) $S(x)$ is. Intuitively, this means that exit times close to x are very similar to a weighted average of exit times close to x when the process jiggles a lot, since it takes the process little time to get to adjacent points.

Discretization. We could now just apply Green's Function and solve the ODE (13) mechanically. But I think it is worth the time to first build some more intuition and to use our discrete-grid approximation to get a better feel of the maths. We maintain the equally-spaced grid with mesh size $\Delta x_n = 1/n$, but to accommodate the time-varying volatility of X_t, we will now change the length of the time increments Δt contingent on the state $\tilde{X}_t$. Specifically, the time interval $\Delta t_n(\tilde{x}_i)$ at grid point $\tilde{x}_i$ should be such that

$$\mathbb{E}[\Delta \tilde{x}_i]^2 = (\Delta x_n)^2 = s(\tilde{x}_i)^2 \Delta t_n(\tilde{x}_i),$$

since the underlying continuous-time diffusion has variance $\mathbb{E}_t\big[s(X_t)\mathrm{d}W_t\big]^2 = s(X_t)^2\mathrm{d}t$ for short time increments. Solving the last equation for $\Delta t_n(\tilde{x})$ gives us

$$\Delta t_n(\tilde{x}) = \frac{(\Delta x_n)^2}{s(\tilde{x})^2} = \frac{1}{n^2 s^2(\tilde{x})}.$$

We see that we have to choose time increments shorter the higher the volatility is at a point $\tilde{x}$. This makes sense; after all, uncertainty is resolved earlier the

higher the volatility is. We see the close correspondence that the varying time increments $\Delta t_n(\tilde{x})$ have to the change of the time measure that we constructed for the continuous process: Recall that $\mathrm{d}\tau = m(X_t)\mathrm{d}t = \mathrm{d}t/s^2(x)$. Note also that on the discrete grid, the subjective time measure $\tau(\cdot)$ corresponds to the cumulative number of steps, N, that the process has taken up to t. If the process visits areas of high volatility, many steps in a short time interval, whereas fewer steps are taken in low-volatility regions.

With this in mind, we can now make the connection from time-varying volatility to the standard case with constant volatility: We had already found a formula for the expected number of times the process will visit a grid point $\tilde{x}_j$ given that it is started at a point $\tilde{x}_i$. This number, $N(\tilde{x}_i, \tilde{x}_j)$, was independent of the size of the time increments and is thus also valid now. To obtain the expected exit time, we just have to multiply this number by the time interval the process stays at grid point $\tilde{x}_j$ each time it gets there, and we get expected local time:

$$F(\tilde{x}_i) = \sum_{j=1}^{n-1} N(\tilde{x}_i, \tilde{x}_j)\Delta t_n(\tilde{x}_j) f(\tilde{x}_j) = \sum_{j=1}^{n-1} \underbrace{\frac{2E_i(\tilde{x}_j)}{s^2(\tilde{x}_j)}}_{\text{expected local time at } \tilde{x}_j} f(\tilde{x}_j).$$

As expected, expected local time depends inversely on the volatility $s^2(\tilde{x})$ at a grid point $\tilde{x}$.

Back to the continuum. Now that we have thoroughly understood, let's go back and apply Green's Function to solve the ODE (13). It turns out that the solution is the exact analog to the equation that we have derived in the discrete world:

$$F(x) = \int_0^1 \underbrace{\frac{2G(x,y)}{s^2(y)}}_{\text{expected local time at } y} f(y)\mathrm{d}y; \tag{14}$$

recall that the Green Function, $G(\cdot)$ and the expected number of visits, $N(\cdot)$, had exactly the same form. As is intuitive, regions with low volatility have higher expected local time since the process is trapped in these for longer.

Expected exit time. Again, calculating the expected exit time from the interval given X_0, $S(X_0)$, just requires us to to set $f(x) = 1$ in the previous equation. Doing this, we obtain

$$S(x) = \int_0^1 \frac{2G(x,y)}{s(y)^2}\mathrm{d}y. \tag{15}$$

Since the Green Function puts most weight on the area where the process starts (it has its peak at X_0), our intuition is confirmed that the expected exit time will be higher when we start the process in a low-volatility region.

Exit probabilities. The probability that the exit occurs at the top of the interval, $P(X_0)$, is even easier to find. Go again over the steps that we took for the constant-volatility case, $s(x) = \sigma$, and you see that our argument for $P(\cdot)$

is unaffected by the volatility of the process – P should still have a constant slope, i.e. $P''(x) = 0$ is the ODE to solve. So the formula

$$P(x) = x \tag{16}$$

we had found in Eq. (3) is still valid. Intuitively, a high volatility $s(x)$ at a point x just means that uncertainty is resolved faster at that point, but the probability of going up or down is not affected, which is what ultimately matters for $P(\cdot)$. Again, the approximation by the discretized process is in line with this intuition.

1.8 Adding drift: Warping space

Adding drift. In order for our results to be applicable to *any* diffusion process, we still have to deal with drifts. For simplicity, we first add a constant drift to a diffusion with constant volatility:[13]

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t,$$

where μ is the drift parameter. The left panel of Figure 5 shows three sample paths of this process for parameters $\mu = \sigma = 1$ and starting value $X_0 = 0$. We see that in general, the realizations exhibit an upward tendency; however, the noise component is strong enough for one of the paths (the blue one) not to stray far from zero. Just as we warped the *horizontal* axis (time) in the previous section, we will now try to warp the *vertical* axis (space). Our goal is to convert the process into a martingale, to which we can then apply the formulas we already know.

Finding the space-warp transform. Formally, we aim to find a function $v(x)$ to apply to the space dimension, x. $v(\cdot)$ should be monotone increasing: We want the warped space to have the same notion of up and down as the original space. As you can guess, we will have to shrink distances for high values of X_t but stretch out distances for low realizations so that our new process $U_t \equiv v(X_t)$ has mean-zero increments (and is thus a martingale); you can see this idea in the right panel of Figure 5.

We can find the exact properties that the function $v(\cdot)$ has to satisfy by applying the Ito Formula to the transform $v(X_t)$ of the original process X_t:

$$\mathrm{d}U_t = \mathrm{d}v(X_t) = v'(X_t)\big[\mu \mathrm{d}t + \sigma \mathrm{d}W_t\big] + \tfrac{1}{2} v''(X_t)\sigma^2 \mathrm{d}t.$$

Since we want U_t to be a martingale, we now impose $\mathbb{E}_t[\mathrm{d}U_t] = 0$. Since $\mathbb{E}_t[\mathrm{d}W_t] = 0$, this yields the following restriction on the function $v(\cdot)$:

$$v''(x) = -\underbrace{\frac{2\mu}{\sigma^2}}_{\equiv \tilde{\mu}} v'(x) = -\tilde{\mu} v'(x). \tag{17}$$

[13] We will make the drift a function of the state in the next section and give formulas for this general case there. We now focus on this special case because it has simple closed-form solutions and gives us a good intuition.

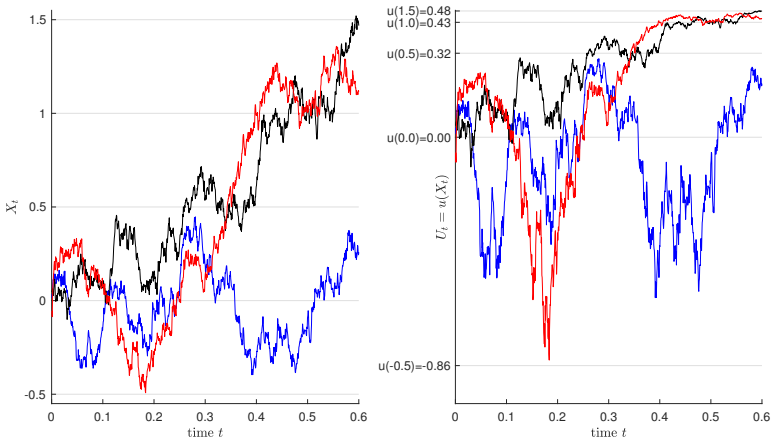


Figure 5: Warping space for the process $\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t$ with $\mu = \sigma = 1$

As anticipated, the function $v(\cdot)$ will have to be concave if μ is positive, since σ^2 and $v'(x)$ are positive. If μ (and thus $\tilde{\mu}$) is negative, on the other hand, we have to make $v(\cdot)$ convex.[14] The stronger the drift μ, the more curvature we have to give to the function $v(\cdot)$, i.e. the more we have to bend space to convert X_t to a martingale.

We should read Eq. (17) as a differential equation for $v'(\cdot)$: It tells us how the first derivative of $v(\cdot)$ should change as we increase x. Since there are no other restrictions than (17) on $v(\cdot)$, there are many functions that satisfy Eq. (17). Mathematically, if some function $v(\cdot)$ satisfies (17), then also any affine transform, i.e. any function $\tilde{v}(x) = \alpha + \beta v(x)$ with arbitrary coefficients $\alpha, \beta \in \mathbb{R}$ and $\beta \neq 0$, will solve (17). Intuitively, there are two degrees of freedom: It is arbitrary which point in space we call zero, and it is also arbitrary how we measure distances in the new space (say in meters, centimeters, or millimeters).

To keep things simple, let us choose the normalizations $v(0) = 0$ and $v'(0) = 1$, meaning that zero is the same point in both the original and the warped space and that we measure distances in the neighborhood of zero the same way in both spaces. With this in hand, we now go back to the task of solving the ODE (17). When dividing by $v'(x)$, the ODE (17) tells us that $v'(x)$ should grow at rate $-\tilde{\mu}$. Since the exponential function is constructed to give us constant growth, we can solve for $v'(x)$ starting at $x = 0$ to find

$$v'(x) = e^{-\tilde{\mu}x} v'(0) = e^{-\tilde{\mu}x}, \tag{18}$$

where the second step follows from our normalization $v'(0) = 1$. To find $v(\cdot)$, finally, we have to integrate up the increments in $v(\cdot)$ from $v(0) = 0$ on, i.e. we

[14] For economists: This differential equation for $v(\cdot)$ should remind you of constant absolute risk aversion.

have to solve $v(x) = v(0) + \int_0^x v'(y)\mathrm{d}y$. This integral has a closed-form solution and our space-warp function $v(\cdot)$ thus reads – since we normalized $v(0) = 0$ –

$$v(x) = \frac{1 - e^{-\tilde{\mu}x}}{\tilde{\mu}}. \tag{19}$$

This equation tells us that we have to bend space so strongly that the resulting process U_t has an upper bound: We have $\lim_{x\to\infty} = 1/\tilde{\mu}$.

This is clearly visible in the right panel of Figure 5, which shows the three realizations of the process $U_t = v(X_t)$ that are obtained by applying the transform $v(\cdot)$ to the original sample paths. We can see that we have managed to make U_t drift-free, albeit at the cost of making volatility shrink for higher values of U_t. We observe that negative realizations of the process X_t bring down U_t a lot, which is outweighed by the fact that the majority of sample paths ends up above the initial value $v(X_0) = v(0) = 0$; note that this must be true since the original process X_t had positive drift! So the median of U_t is increasing in t, but the mean $\mathbb{E}_0[U_t]$ is zero. In the language of stochastic processes, we say that we have brought the process X_t to *natural scale* by applying $v(\cdot)$.

Discrete approximation. How can we think about the warping of space in the discrete world? We will maintain the binomial structure of the process (the fact that it either goes up one point with probability 50% or down one point with probability 50% at each grid point), which will require us to adapt the spacing of the grid points. Let us start by fixing the first two grid points, $\tilde{x}_0 = 0$ and $\tilde{x}_1 = \Delta x_1$ for some $\Delta x_1 > 0$; this step amounts to the normalization $v(0) = 0$ and $v'(0) = 1$ that we carried out in the continuous case. We then find the grid points $\{\tilde{x}_n\}_{n=2}^N$ and the time increments $\{\Delta t_n\}_{n=1}^{N-1}$ inductively by making the conditional mean and the conditional variance of the discretized process match those of the continuous process, i.e. we require

$$\begin{aligned}
&\mathbb{E}_t[\tilde{X}_{t+\Delta t_n}] - \tilde{X}_t = \tfrac{1}{2}(\tilde{x}_{n+1} - \tilde{x}_n) + \tfrac{1}{2}(\tilde{x}_n - \tilde{x}_{n-1}) = \mu\Delta t_n, \\
&\mathbb{E}_t[\tilde{X}_{t+\Delta t_n} - (\tilde{X}_t + \mu\Delta t_n)]^2 = \\
&\qquad\qquad = \tfrac{1}{2}(\tilde{x}_{n+1} - \tilde{x}_n + \mu\Delta t_n)^2 + \tfrac{1}{2}(\tilde{x}_{n-1} - \tilde{x}_n + \mu\Delta t_n)^2 = \sigma^2\Delta t_n.
\end{aligned}$$

This is somewhat cumbersome to solve; however, it is clear what kind of grid we would obtain for an upward-drifting process: The step we take upward must be larger than the step we take downward. The space transform is then to simply set $v(\tilde{x}_i) = i$ and bring the process to a linear grid. We clearly see that the unequal spacing of grid points performs the same transform as the function $v(\cdot)$ in the continuous case: For a process with upward drift, the space transform shrinks distances for high values of the process in both cases.

Exit probabilities. We can now go ahead and apply the formulas we found for martingales to our modified process, U_t. The transformed process $U_t = v(X_t)$ is a martingale that lives on the interval $[v(a), v(b)]$ and has initial value $U_0 = v(X_0)$. For the probability that X_t exits at the upper boundary b, we can use Eq. (10) to find

$$P_{ab}(X_0) = P_{v(a)v(b)}\big(v(X_0)\big) = \frac{v(X_0) - v(a)}{v(a) - v(b)}, \tag{20}$$

where the space transform $v(\cdot)$ was given in Eq. (19).

Expected local time and expected exit time. To obtain expected local time, $G_{ab}(X_0, x)$, and the expected exit time, $S(X_0)$, we have to take one more step: We have to find the volatility of the warped process U_t. The Ito Formula tells us that

$$\mathrm{d}U_t = \mathrm{d}v(X_t) = \underbrace{v'(X_t)s(X_t)}_{=s_u(X_t)}\mathrm{d}W_t = e^{-\tilde{\mu}X_t}\sigma\mathrm{d}W_t.$$

When applying the Ito Formula, the terms $v'(X_t)\mu\mathrm{d}t$ and $\frac{1}{2}v''(X_t)s(X_t)^2\mathrm{d}t$ cancel since we constructed $v(\cdot)$ precisely to *make* them cancel. We find that the volatility of the process U_t is given by $s_u(X_t) = \exp(-2\tilde{\mu}X_t)\sigma$, when expressed in terms of X_t. We can now use the formula (14) to calculate expectations over any function $\hat{f}(U_t)$, since U_t is a martingale:

$$\hat{F}(U_0) = \mathbb{E}_0\int_0^T \hat{f}(U_t)\mathrm{d}t = \int_{v(a)}^{v(b)} \frac{2G_{v(a)v(b)}(U_0, u)}{s_u^2\big(v^{-1}(u)\big)}\hat{f}(u)\mathrm{d}u,$$

where we have adapted the formula to deal with an interval of arbitrary size (see also the following section). We now go back from U-space to X-space; this change of variable requires us to adapt the integral according to $\mathrm{d}u = v'(x)\mathrm{d}x$. Using the definition of $s_u(\cdot)$ this yields

$$F(X_0) = \int_a^b \underbrace{\frac{2G_{ab}(X_0, x)}{\sigma^2 e^{-\tilde{\mu}x}}}_{\text{expected local time at } x} f(x)\mathrm{d}x,$$

where we define $f(x) \equiv \hat{f}(v(x))$ – since $\hat{f}(\cdot)$ was an arbitrary function, we may as well chose this definition. We see that expected local time depends again inversely on volatility σ^2. The new term is the division by $v'(x) = e^{-\tilde{\mu}x}$: This term assigns more local time to high x-values for a process with positive drift; this makes sense, since the process is expected to spend more time at high x-values if $\mu > 0$.

Again, the expected exit time can be found by setting $f(x) = 1$. Since the Green Function has two linear parts (to the left and the right of X_0), it is possible to find a closed-form solution for $S_{ab}(X_0)$ since the integral $\int_c^d xe^{\alpha x}\mathrm{d}x$ allows for a closed-form solution; we omit this here for laziness reasons.

Exercise: We can also calculate expeceted local time in the U-world instead of the X-world. To do this, first find the function $\hat{s}_u(u)$ in the SDE

$$\mathrm{d}U_t = \hat{s}_u(U_t)\mathrm{d}W_t.$$

Then calculate expected local time for all $u \in \big(v(a), v(b)\big)$ and compare to local time at $x = v^{-1}(u)$. Explain.

1.9 The general case: Some formulas

Consider now a general diffusion process

$$\mathrm{d}X_t = m(X_t)\mathrm{d}t + s(X_t)\mathrm{d}W_t$$

on the interval (a, b). Here, $m(\cdot)$ is the drift and $s(\cdot)$ is the volatility of the process. We assume again that the function $s(\cdot)$ is bounded away from zero, i.e. that $s(x) \geq \epsilon$ for all $x \in (a, b)$ some $\epsilon > 0$

The space transform. We first warp space, which will already be enough to obtain the exit-probability function $P(\cdot)$. Following the same steps as in the previous section, we find that the monotone transform $v(\cdot)$ has to satisfy

$$v''(x) = -\frac{2m(x)}{s^2(x)}v'(x),$$

where $s^2(x)$ is short-hand for writing $[s(x)]^2$. Imposing the normalization $v'(q) = 0$ at an arbitrary point $q \in (a, b)$ that we pick from our interval, solving the ODE for $v'(x)$ then gives us

$$v'(x) = \exp\Big(-\int_q^x \frac{2m(y)}{s(y)^2}\mathrm{d}y\Big). \tag{21}$$

You can verify that this is the solution to the above ODE by taking the derivative with respect to x (use the Leibniz Rule). Normalizing again and setting $v(q) = 0$, by integrating over x we find that the space transform is

$$v(x) = \int_q^x \left[\exp\Big(-\int_q^y \frac{2m(z)}{s^2(z)}\mathrm{d}z\Big)\right]\mathrm{d}y. \tag{22}$$

Exit probability. Now, since $U_t = v(X_t)$ is a martingale on the interval $(v(a), v(b))$, the probability that the process exits at the top is given by

$$P_{ab}(x) = \frac{v(x) - v(a)}{v(b) - v(a)}. \tag{23}$$

Expected local time. To find expected local time and the expected exit time, we now derive a SDE for the process U_t. Since we defined $U_t = v(X_t)$, the Ito Formula tells us that

$$\mathrm{d}U_t = \underbrace{v'(X_t)s(X_t)}_{=s_u(X_t)}\mathrm{d}W_t. \tag{24}$$

Since U_t is a martingale, we can now use the formula (14) to calculate expectations over any function $\hat{f}(U_t)$:

$$\hat{F}(U_0) = \mathbb{E}_0\int_0^T \hat{f}(U_t)\mathrm{d}t = \int_{v(a)}^{v(b)} \frac{2G_{v(a)v(b)}(U_0, u)}{s_u^2\big(v^{-1}(u)\big)}\hat{f}(u)\mathrm{d}u, \tag{25}$$

Changing the integration variable from u to x, which requires us to set $\mathrm{d}u = v'(x)\mathrm{d}x$, and using the definition of $s_u(\cdot)$ yields

$$F(X_0) = \int_a^b \underbrace{\frac{2G_{v(a)v(b)}\big(v(X_0), v(x)\big)}{v'(x)s^2(x)}}_{\text{expected local time at } x} f(x)\mathrm{d}x, \tag{26}$$

where $v'(x)$ is given in Eq. (21) and where the Green Function $G_{\cdot\cdot}(\cdot)$ is given in Eq. (11). We have also defined $f(x) = \hat{f}(v(x))$, which is again an arbitrary function. The expression involving the Green Function in (26) looks complicated, but is not not all that cumbersome to compute in applications; after all, the Green Function is piecewise linear.

Expected exit time. As always, we can set $f(x) = 1$ in Eq. (26) to obtain the expected exit time $S_{ab}(X_0)$ from the interval (a, b) as

$$S_{ab}(X_0) = \int_a^b \frac{2G_{v(a)v(b)}\big(v(X_0), v(x)\big)}{s^2(x)v'(x)}\mathrm{d}x, \tag{27}$$

The time transform. In Eq. (26), we recognize the change of the time measure; it takes the form $\mathrm{d}\tau = m(X_t)\mathrm{d}t = s^{-2}(X_t)\mathrm{d}t$ here.

2 Is a boundary ever reached?

In many cases, the integrals that have to be evaluated in equations such as (26) can be very complicated; it is then often easier in practice to compute expected exit times, exit probabilities etc. numerically. However, there is one case in which the analytic approach offers large advantages – and is indeed indispensable: When the volatility function, $s(x)$, approaches zero at one of the interval bounds a and b. In this case, it is usually not clear if the process can reach a boundary with positive probability (in which case we call the boundary *accessible*) or if the boundary is reached with probability zero (an *inaccessible* boundary).

2.1 Geometric Brownian Motion: Explicit (*strong*) solutions to SDEs

To see where the problem lies, let us first consider the important example of *Geometric Brownian Motion*. This process is defined by the SDE

$$\mathrm{d}X_t = \mu X_t\mathrm{d}t + \sigma X_t\mathrm{d}W_t \qquad \text{with } X_0 > 0, \tag{28}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are parameters. The question is now: Can this process ever reach zero, or even go below zero?

Re-writing the SDE (28) as follows suggests it cannot:

$$\frac{\mathrm{d}X_t}{X_t} = \mu\mathrm{d}t + \sigma\mathrm{d}W_t.$$

This equation says that the percentage growth rate of the process (per unit of time) has mean μ and standard deviation σ. One may argue that something that grows or shrinks at a certain rate (even if that rate is random), should always stay positive...

However, we would obtain a different answer from the following thought experiment: If we simulated the SDE (28) on a discrete grid $\{0, \Delta x, 2\Delta x, \dots\}$ following the recipe laid out in previous sections, the algorithm would tell us that at the grid point $\tilde{x}_1 = \Delta x$ there is always a *positive* (if small) probability of ending up at zero; after hitting zero, the process would then stay there forever, for whatever small mesh size Δx.

Which of the two intuitions is right? It turns out that the first one is. We will now see how stochastic calculus can easily settle the issue. Let's consider the process

$$\mathrm{d}Z_t = (\mu - \tfrac{\sigma^2}{2})\mathrm{d}t + \sigma \mathrm{d}W_t \qquad \text{with } Z_0 = \ln X_0,$$

where W_t is the same Wiener Process as the one driving X_t in (28). Now, let us define $\hat{X}_t \equiv \exp(Z_t)$ as a function of that process. Using the Ito Rule, we have

$$\begin{aligned}\mathrm{d}\hat{X}_t &= (\mu - \tfrac{\sigma^2}{2})\hat{X}_t\mathrm{d}t + \tfrac{\sigma^2}{2}\hat{X}_t\mathrm{d}t + \hat{X}_t\sigma\mathrm{d}W_t \\ &= \mu\hat{X}_t\mathrm{d}t + \sigma\hat{X}_t\mathrm{d}W_t.\end{aligned}$$

By this, we have shown that the processes $\hat{X}_t$ and X_t are identical: They have identical increments, $\mathrm{d}\hat{X}_t = \mathrm{d}X_t$, and we had specified the initial condition to be the same, $\hat{X}_0 = \exp(Z_0) = X_0$. Indeed, what we have done is that we have found the following solution to the SDE (28):

$$X_t = X_0 \exp\big((\mu - \tfrac{\sigma^2}{2})t + \sigma W_t\big).$$

This is what is called a *strong solution* to the SDE (28): It tells us how to calculate X_t for any given realization of the Brownian Motion, $\{W_s\}_{s=0}^{\infty}$, that drives the SDE.[15] This is actually the best thing we can possible hope for when it comes to SDEs!

With this solution in hand, we can now come back to our original question: Does X_t ever reach zero? Since the exponential function is always positive, i.e. $X_t = \exp(Z_t) > 0$ for all t, this question is easy to answer: It cannot. We will now see that doing similar things (applying transforms to space, and also time, if necessary) can also settle the question for processes for which we can't solve the SDE explicitly.

2.2 Result for general diffusion processes

For the following class of processes, for example, an explicit solution to the SDE (generically) has not been found:

$$\mathrm{d}X_t = \mu\mathrm{d}t + \sigma X_t^{\alpha}\mathrm{d}W_t \qquad \text{with } X_0 > 0, \tag{29}$$

[15] The other solution concept for SDEs are *weak solutions*; a weak solution is a process that coincides in all its stochastic properties (think: moments) with those implied by the SDE but that is not specified as a function of the Wiener Process that drives the SDE.

where $\alpha \geq 0$, $\mu \in \mathbb{R}$, and $\sigma > 0$. We will study the behavior of this process on the interval (a, b), where $0 = a < b < \infty$.[16]

So do these processes ever reach zero? First, note that just like for Geometric Brownian Motion, the drift and volatility converge to zero as $X_t \to 0$, but does this imply that X_t will not reach zero? If such processes can reach zero, for which values of σ, μ, and α will this occur?

In the following, we will solve this puzzle for two specific cases but also derive results that are valid for a general diffusion process of the form

$$\mathrm{d}X_t = m(X_t)\mathrm{d}t + s(X_t)\mathrm{d}W_t, \qquad X_0 \in (a, b), \tag{30}$$

that lives on an interval (a, b) and for which volatility vanishes at the lower bound of the state space, i.e. for which $\lim_{x \to a} s(x) = 0$. We will follow the same steps as in the sections before: We (i) apply a space transform in order to convert the process into a martingale and (ii) apply a time transform to convert the (space-warped) process into a standard Wiener Process. The following proposition summarizes the results for a general diffusion process. After stating the result, we will only see some heuristic arguments involving discrete approximation processes that will make us confident that these statements hold. For a formal proof, see Helland's (1996) Proposition 3, who in turn takes the proof from the classic probability textbook Breiman (1992).

Proposition 1 (Accessibility of boundaries for one-dimensional diffusions)**.** *Consider the diffusion process given by* (30) *on the interval* (a, b)*, where* $-\infty < a < b < \infty$*. If there exists* $c \in (a, b)$ *for which*

$$\int_a^c \frac{v(x) - v(a)}{v'(x)s^2(x)}\mathrm{d}x < \infty, \tag{31}$$

where $v(x)$ *is given in Eq.* (22) *and* $v'(x)$ *is given in Eq.* (21)*, then*

1. *the expected exit time* $S_{ab}(X_0)$ *from the interval* (a, b) *is given by Eq.* (27)*, and it is finite, i.e.* $S_{ab}(X_0) < \infty$ *for all* $X_0 \in (a, b)$*,*
2. *the boundary* a *is accessible, i.e.* X_t *reaches* a *in finite time with positive probability,*
3. *the probability that* X_t *exits the interval at* a *(instead of* b*) is* $1 - P_{ab}(X_0)$*, where* $P_{ab}(X_0)$ *is given by Eq.* (23)*. Furthermore,* $\lim_{x \downarrow a}[1 - P_{ab}(x)] = 1$*.*

If the integral in (31) *is infinite for all* $c \in (a, b)$*, then*

1. *the expected exit time from* (a, b) *is infinite for any initial value* $X_0 \in (a, b)$*,*

[16] One thing to notice here is that the volatility function $s(x) = \sigma x^\alpha$ is not Lipschitz-continuous at $x = 0$ if $\alpha < 1$. Thus we cannot apply the fundamental existence and uniqueness result that guarantees a solution to our SDE. However, such processes can still be studied by applying time and space transforms, for and no Lipschitz arguments are needed for these arguments to be valid. From a more formal standpoint, this is explained in Helland's (1996) paper, who nicely summarizes results on such processes from an early literature in stochastic calculus (Feller, Ito, McKean, Dynkin, and Breiman).

2. *the boundary a is inaccessible, i.e. the probability that X_t reaches a in finite time is zero.*

Remark: For the upper boundary b, an analogous result holds, but the numerator in the integral in Eq. (31) changes to $v(b) - v(x)$ and the range of integration has to be changed to $[c, b]$.

2.3 Example 1: A driftless diffusion

We will first illustrate Step (ii), which involves only a time transform and study an example that needs both steps (a space and a time transform) later. Let's set $\mu = 0$ in the SDE (29) and study the following class of diffusion processes:

$$\mathrm{d}X_t = \sigma X_t^\alpha \mathrm{d}W_t \qquad \text{with } X_0 \in (0, b),$$

where $\sigma > 0$ and $\alpha > 0$. We already know that for $\alpha = 1$ (a specific case of Geometric Brownian Motion), zero is never reached. In fact, it will turn out that this case ($\alpha = 1$) is the knife-edge case: For all values below, the process reaches zero, whereas it doesn't for any value above.

To see how the test integral in Eq. (31) comes about, let us calculate the expected exit time of our process from the interval $[0, b]$. For lower boundaries at positive values, $a > 0$, we already know how to compute the expected exit time; let's now just confidently take the limit as $a \to 0$. We don't need a space transform to have a martingale, i.e. we can set $v(x) = x$ in Eq. (26) and $f(x) = 1$ and obtain the expected exit time as

$$S_{0b}(X_0) = \lim_{a \downarrow 0} S_{ab}(X_0) = \int_0^b \frac{2G_{0b}(X_0, x)}{\sigma^2 x^{2\alpha}} \mathrm{d}x.$$

Now, recall that the Green function is hat-shaped and peaks at X_0; it has two linear pieces: one on $(0, X_0)$ and a second on $[X_0, b)$. The second piece on the interval $[X_0, b)$ doesn't give problems when integrating: Since the variance of the process is bounded away from zero, the integral will always take a finite value on $[X_0, b)$. For the lower part $(0, X_0)$, however, the variance goes to zero and thus the integrand goes to infinity. Since the Green Function is linear, S_{0b} is finite if and only if

$$\int_0^{X_0} \frac{x}{\sigma^2 x^{2\alpha}} < \infty,$$

which is exactly what Eq. (31) in the proposition says. Thus, we can conclude that for the test integral in the proposition we have

$$\int_0^c \frac{x}{\sigma^2 x^{2\alpha}} \mathrm{d}x = \frac{1}{\sigma^2} \int_0^c x^{1-2\alpha} \mathrm{d}x \begin{cases} = \infty & \text{if } \alpha \geq 1, \\ < \infty & \text{if } \alpha \in [0, 1). \end{cases}$$

We summarize the result as follows:

- $\alpha \in (0, 1)$: The boundary $a = 0$ is accessible.

- $\alpha \geq 1$: The boundary $a = 0$ is inaccessible.

It turns out that again our discrete approximating process will give us a good intuition for the result. At first, this seems odd: After all, as we saw before, any approximating process reaches the grid point $\tilde{x}_0 = 0$, with positive probability in finite time. So how can it be that in the limit the process does not reach zero? The key will be time: It will take infinite time to get there for some values of α.

To see when this happens, it turns out that it is instructive to consider only how long it takes to exit the interval from the first grid point, $\tilde{x}_1$, of a grid with mesh size $\Delta x_n = 1/n$. The time increment Δt_n at this first grid point is

$$\Delta t_n(\tilde{x}_1) = \frac{\Delta x^2}{s^2(\tilde{x}_i)} = \frac{n^{-2}}{\sigma^2 n^{-2\alpha}} = \frac{n^{2(\alpha-1)}}{\sigma^2}.$$

We see that $\Delta t_n(\tilde{x}_1)$ approaches infinity if $\alpha > 1$ – which is exactly the same threshold value that the test integral in the proposition gave us! Thus, even when drawing only negative shocks in the approximating process, it must be that the time to reach zero goes to infinite as the grid becomes infinitely fine and we must conclude that we cannot reach the point zero in the limit.

On the other hand, if $\alpha \in (0, 1)$ we see that $\Delta_t(\tilde{x}_1)$ goes to zero as the grid becomes infinitely fine. Thus it makes sense what the proposition says: The boundary $a = 0$ stays accessible in the limit (as does the expected exit time), and the probability of exiting zero approaches one on the first grid point as the grid gets finer.

Only for the knife-edge case $\alpha = 1$, the discrete approximating process does not provide us with an answer – but recall that we had settled the question if Geometric Brownian Motion reaches zero by stochastic calculus already.

Now, you may wonder the following: Our time transform generated a new process $U_t = v(X_t)$, which by construction is a standard Wiener Process. But a standard Wiener Process always reaches zero with positive probability – so how can that be compatible with X_t never reaching zero? Figure 6 shows what happens: Time actually runs out for the transformed processes! You see that when X_t is plotted against subjective time, $\tau(t)$, the processes that approach zero just run dry. Geometrically, the function $\tau(t)$ would flatten out and asymptotes to a finite value: As weird as this may seem, the traveler would feel subjective time to be finite, while true time runs on forever. Again, this is in line what happens for the approximating discrete processes – think about it!

2.4 Example 2: A process with drift

Let us now study an example with drift. Let us set $\alpha = \frac{1}{2}$ in Eq. (29) and consider the diffusion process

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t,$$

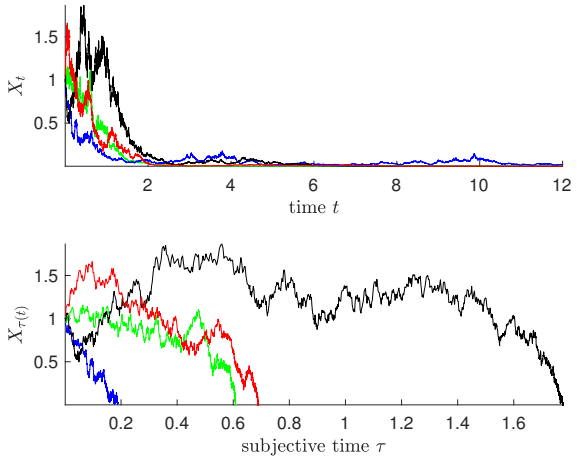


Figure 6: Processes running dry. Sample paths of diffusion $\mathrm{d}X_t = X_t \mathrm{d}W_t$ with negative shock realizations.

where $\mu, \sigma \in \mathbb{R}$. For the case without drift ($\mu = 0$), this process is called the *Feller diffusion*. Our results from the previous subsection tell us that the process with $\mu = 0$ will hit zero, but it turns out that this need not be the case for positive drifts.

Since the process has a drift, we first have to apply a space transform to convert it a martingales. Let us normalize the slope of the transform to unity at $x = 1$, i.e. let us choose $q = 1$ in Eq. (21); the slope of the space transform then has to be

$$\begin{aligned} v'(x) &= \exp\Big(- \int_1^x \frac{2\mu}{\sigma^2 y} \mathrm{d}y \Big) = \exp\big(- \tfrac{2\mu}{\sigma^2} [\ln y]_1^x \big) \\ &= \exp\big(- \tfrac{2\mu}{\sigma^2} \ln x \big) = x^{-\frac{2\mu}{\sigma^2}} = x^{-\tilde{\mu}}, \end{aligned}$$

where we have defined $\tilde{\mu} \equiv 2\mu/\sigma^2$. As we know from before, the space transform v is concave for positive drifts and convex for negative drifts. Again normalizing $v(1) = 0$, by integration we find the space transform itself to be

$$v(x) = \int_1^x v'(y) \mathrm{d}y = \begin{cases} \frac{x^{1-\tilde{\mu}} - 1}{1 - \tilde{\mu}} & \text{if } \tilde{\mu} \neq 1 \text{ (i.e. if } \mu \neq \frac{\sigma^2}{2}), \\ \ln x & \text{if } \tilde{\mu} = 1 \text{ (i.e. if } \mu = \frac{\sigma^2}{2}), \end{cases}$$

which should remind the economists among us of CRRA preferences. We see that for $\tilde{\mu} \geq 1$, space is bent so strongly that $v(x)$ approaches minus infinity as $x \to 0$. The warped process $U_t \equiv v(X_t)$ will never reach $-\infty$, thus we expect that X_t never reaches zero (under this parameter configuration). Indeed, when

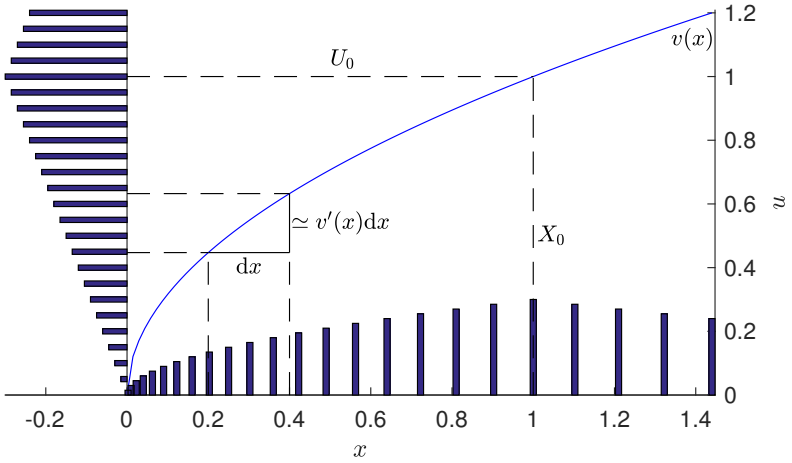
calculating the test integral in (31) in the proposition, we obtain

$$\int_{v(0)}^{v(b)} \frac{x^{1-\tilde{\mu}} - 0}{(1-\tilde{\mu})(\sigma\sqrt{x})^2 x^{-\tilde{\mu}}} \mathrm{d}x = \frac{1}{(1-\tilde{\mu})\sigma^2} \int_{v(0)}^{v(b)} 1 \mathrm{d}x \qquad \begin{cases} < \infty & \text{if } \tilde{\mu} < 1, \\ = \infty & \text{if } \tilde{\mu} \geq 1, \end{cases}$$

since the range of integration becomes infinite if $\tilde{\mu} \geq 1$. We summarize our result as follows:

- $\mu < \frac{\sigma^2}{2}$: Boundary $a = 0$ is accessible.
- $\mu \geq \frac{\sigma^2}{2}$: Boundary $a = 0$ is inaccessible.

Let us use again the intuition from a discrete approximating process to get some more intuition. Figure 7 visualizes how the test integral in the proposition comes about. The curved line shows the space transform from x (on the horizontal axis) to u (on the vertical axis). For the chosen parameters, $\mu = \frac{1}{4}$ and $\sigma = 1$, the transform takes the form of the square-root function. Recall that the space transform is such that the process $U_t = v(X_t)$ is a martingale. As we saw in previous sections, we can approximate any martingale by a binomial process on an evenly spaced grid, making sure that we adjust the time increments in the right fashion to take into account state-varying volatility.



Stochastic Process $\mathrm{d}X_t = \mu \mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t$ with $\sigma = 1$, $\mu = \frac{1}{4}$, and $X_0 = 1$ on the interval $(a, b) = (0, 2)$. Space transform: $v(x) = x^{1-\tilde{\mu}} = \sqrt{x}$, where $\tilde{\mu} = \mu\sigma^2/2$. Bars show expected number of visits of a discretized approximating process before exiting (a, b).

Figure 7: Space transform and expected number of visits for discretized process

Now, the bars on the u-axis show $N(\tilde{U}_0, \tilde{u}_i,)$, the expected number of times that the approximating process $\tilde{U}_t$ will visit a grid point $\tilde{u}_i$ before exiting the interval $(0, b)$ and when started at $\tilde{U}_0 = 1$. You surely recall that $N(\cdot, \cdot)$ has the same hat shape as the Green Function has (see Figure 3). When we translate

the approximating process $\tilde{U}_t$ back to the x-axis using the transform $v^{-1}(\tilde{u}_i)$ to obtain the grid points $\{\tilde{x}_i\}$, we obtain the unevenly-spaced that you see marked by the bars on the x-axis. Note that the expected number of visit times for a grid point $\tilde{x}_i$ is the same for the corresponding grid point $\tilde{u}_i = v(\tilde{x}_i)$. We note that for the x-grid, the spacing between the bars increases in x in order to accommodate the positive drift.

Finally, note that we can visualize the test integral in (31) as the sum over the bars on the x-axis times $\Delta t_n(\tilde{x})$, the time that the process stays at a grid point, which again is inversely proportional to $s^2(x)$. Note that the division by $v'(x)$ takes care of the change of variable that $v^{-1}(\cdot)$ performs: The number of bars on a given interval $\mathrm{d}x$ on the x-axis is inversely proportional to $v'(x)$. In the example depicted in the graph, the limit of these weighted sums will be finite; however, this is not the case any more for high-enough drifts since the spacing of grid points at zero becomes so dense that the limit diverges to infinity.

3 Forward-iterating densities

Often, we are also interested in obtaining a probability distribution, or density, for a stochastic process that exits an interval at the boundaries of an interval. In our application to firms, for example, we may be interested in the distribution of firms over productivity states. For one important case, there are closed-form solutions that are worth knowing: the case of a random walk with constant variance and constant drift on an interval with *one* finite boundary. This solution was exploited, for example, in Luttmer (2007).[17]

So let us consider now processes on an interval (a, b) where a is a finite number and where $b = \infty$. In the firm example, let us interpret the process X_t hitting a as the firm exiting the market. The stochastic process is given by

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t, \qquad \text{where } X_0 > a$$

and where we assume $\sigma > 0$ and $\mu \in \mathbb{R}$. For each interior point $x \in (a, \infty)$, the density $g(x, t)$ of firms that have not exited until time t must solve the following *Kolmogorov Forward Equation* (see my lecture notes on the Forward Equation for how this PDE is obtained):

$$g_t(x, t|X_0) = -\mu g_x(x, t) + \tfrac{\sigma^2}{2} g_{xx}(x, t), \tag{32}$$

where subscripts to g denote partial derivatives. The boundary condition for $g(\cdot)$ at the lower interval end is

$$g(a, t|X_0) = 0 \qquad \text{for all } t \in (0, \infty), \tag{33}$$

Where does this boundary condition come from? Intuitively, the jittering of the Brownian Motion is so strong that close to the boundary exit is almost certain.

[17] In the case of two boundaries, closed-form solutions sometimes exist, but they are so complicated that a numerical solution is preferable in my view.

Indeed, we can see this in our equations for the expected exit time and the exit probability as x that we have derived before: If $\sigma > 0$, then the expected exit time $S(x)$ tends to zero as $x \to a$; at the same time, the probability $P(x)$ for the process to exit the interval (above) at b tends to zero. We can also derive the boundary condition from a discrete approximating process. If you do the maths, you will see that the density $\tilde{g}(\tilde{x}_1, t)$ at the first grid point $\tilde{x}_1$ above a is (roughly) a local average of the densities at $t - \Delta t_n$ at the neighboring grid points, $\tilde{g}(\tilde{x}_0, t - \Delta t_n)$ and $\tilde{g}(\tilde{x}_2, t - \Delta t_n)$. Since the density at a is always zero – these firms have exited –, the probability $\tilde{g}(\tilde{x}_1, t)$ is roughly the average between zero and another number. Thus, in the limit as the grid becomes finer and as $\Delta x_n \to 0$, we must have $\tilde{g}(\tilde{x}_1, t) \to 0$ as $n \to \infty$.

It turns out that the the closed-form solution to the PDE (32) with boundary condition (33) is the following mixture of two normal distributions:

$$g(x, t|X_0) = \underbrace{\frac{1}{\sigma\sqrt{t}}\phi\left(\frac{x - X_0 - \mu t}{\sigma\sqrt{t}}\right)}_{=n_A(x,t)} \\ \underbrace{- \frac{1}{\sigma\sqrt{t}} \exp\left(-\frac{2\mu(X_0 - a)}{\sigma^2}\right)\phi\left(\frac{x + X_0 - 2a - \mu t}{\sigma\sqrt{t}}\right)}_{=n_B(x,t)},$$

where $\phi(\cdot)$ is the standard normal density function. Figure 8 breaks up the solution into the two components $n_A(\cdot)$ and $n_B(\cdot)$ and plots the solution for different times t.

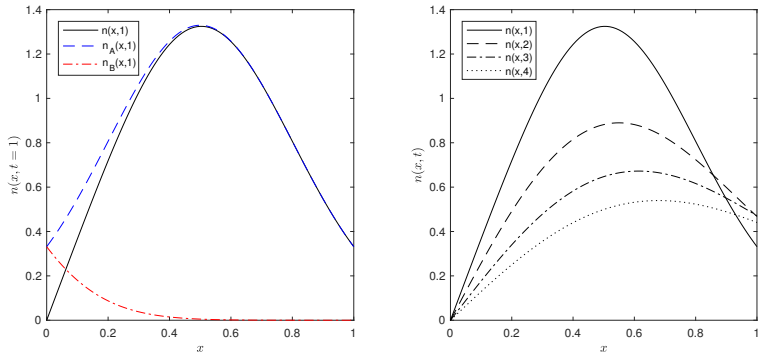


Figure 8: Density of process over time with boundary $a = 0$ and starting value $X_0 = \frac{1}{2}$ ($\sigma = 0.3$, $\mu = 0$, $b = \infty$)

When starring at the formula for $g(\cdot)$, we note the following. First, as t goes to zero, the density collapses to a point mass $\delta_{X_0}(x)$ at the initial value X_0. Second, for small values of t, the term $n_A(x, t)$ dominates. This term describes a normal distribution centered around $X_0 + \mu t$ with standard deviation $\sigma\sqrt{t}$;

this is nothing but the distribution of a random walk with drift μ and variance σ^2. As t grows larger, the term $n_B(x,t)$ grows in importance; this second term corrects for the mass of paths that have exited the interval until t and ensures that the boundary condition $n_B(a,t) = 0$ is met.

The second term also contains a normal density – but where on earth does this second normal distribution come from? To get an idea, let us write out the density for an especially simple case: Let's consider a random walk without drift and standardized variance, i.e. let's set $\mu = 0$. Also, let us set the lower boundary to $a = 0$. Then, for this special case, we can re-write the density as

$$\hat{g}(x,t|X_0) = \frac{1}{\sigma\sqrt{t}}\phi\left(\frac{x - X_0}{\sigma\sqrt{t}}\right) - \frac{1}{\sigma\sqrt{t}}\phi\left(\frac{x - (-X_0)}{\sigma\sqrt{t}}\right).$$

We see that the second term is a standard normal density with variance $\sigma^2 t$ that is centered around $-X_0$, whereas the first term is centered around X_0. This may seem somewhat esoteric at first, but it can be made sense of. We can think about this second term as a *mirror source* of paths that is just as far away from the boundary $a = 0$ as X_0 is, but that lies on the other side. Figure 9 shows how this works. There are three types of paths: The first type (drawn in black) never touches the boundary a until $t = 1.5$; the second type of paths (drawn in red) crosses the boundary at least once but ends up above $a = 0$ at $t = 1.5$; and a third type (drawn in blue) which crosses the boundary at least once and ends up below $a = 0$ at $t = 1.5$. Obviously, only paths of Type 1 should be counted in our density.

Now, let us see where the mirror source comes in. For each path, there is a corresponding "mirrored" sister path coming from the mirror source $\tilde{X}_0 = -1$, which are shown in the figure as dashed lines. We note that the term $n_A(\cdot)$ in our solution for the density counts paths of Types 1 and 2, but not those of Type 3. It is fine that we count paths of Type 1, but we have to correct for those of Type 2, which should be removed. The term $n_B(\cdot)$ does this, but it actually subtracts the mass of Type-3 mirror paths, not that of Type 2.

Why is this correct? Note that for any path of Type 2, there exists exactly on path of Type 3 that turns around at one of the crossings of the boundary a and occurs with the same likelihood; intuitively, the red dashed path could just as well have stayed above zero and followed the red solid line as going back below zero. Think about a discrete approximating process that goes up instead of down when it first crosses the boundary to see that this must be true.

Reflected Brownian Motion. It turns out that the mirror-source trick also helps us to construct the density for another class of problems. Let us suppose that paths that touch the bound a are not absorbed, but are reflected upward again. I am not aware of any economic application where this is a reasonable assumption, but never mind... In our example with $a = 0$, this gives rise to the modified process $X_t^{(r)} = |X_t|$. To visualize the sample paths of this process, just replace the solid paths in Figure 9 by the dashed paths of the same color whenever the solid path is below $a = 0$.

To obtain the density of the reflected process $X_t^{(r)}$, it turns out that we have

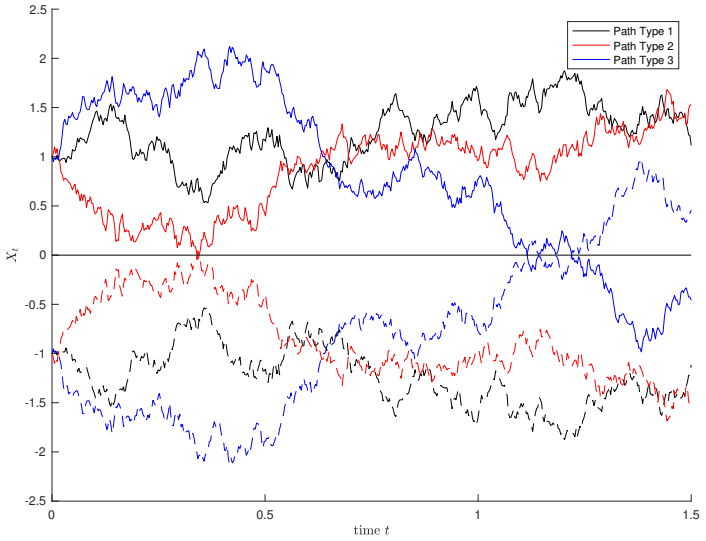


Figure 9: Process $\mathrm{d}X_t = \sigma \mathrm{d}W_t$ with initial condition $X_0 = 1$, $\sigma = 1$, bound $a = 0$, and mirror source at $\hat{X}_0 = -1$

to *add* the term $n_B(\cdot)$ instead of subtracting it and the solution is (assuming again $a = 0$ and $\mu = 0$) as follows:

$$\hat{g}^{(r)}(x, t|X_0) = \frac{1}{\sigma\sqrt{t}}\phi\left(\frac{x - X_0}{\sigma\sqrt{t}}\right) + \frac{1}{\sigma\sqrt{t}}\phi\left(\frac{x - (-X_0)}{\sigma\sqrt{t}}\right).$$

Why is this? The term $n_A(\cdot)$ gives us the correct density for all paths of Types 1 and 2. For each path of Type 3, there exists a sister path emanating from the mirror source that ends up at the same place as the original path does. Thus, we have to add the density of these paths, which is the term in the above formula, $n_B(\cdot)$.

4 Higher dimensions: PDEs for exit times, exit probabilities, and densities

In the case of stochastic processes more than one dimension, I am not aware of closed-form solutions that are helpful in economic applications.[18] In these cases, it is usually more practical to solve for exit times and other variables solving a PDE using a finite-difference method. We will now state these PDEs.

[18] When the set in question is a ball (i.e. a circle in 2 dimensions or a sphere in 3 dimensions), then changing to polar coordinates around the center of the ball can help. This is usually not a set of interest for economic applications, but it can be of great use in physics.

In general, we will consider vector-valued diffusion processes, $\mathbf{X}_t \in \mathbb{R}^m$, defined by the SDE

$$\mathrm{d}\mathbf{X}_t = \mathbf{m}(\mathbf{X}_t)\mathrm{d}t + \mathbf{\Sigma}(\mathbf{X}_t)\mathrm{d}\mathbf{W}_t,$$

where $\mathbf{m} : \mathbb{R}^m \to \mathbb{R}^m$ is the drift vector as a function of the state, $\mathbf{\Sigma} : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ is an m-by-m-matrix-valued function which determines the covariance structure of shocks, which may depend on the state, and where $\mathbf{W}_t$ is an m-dimensional Wiener Process. The process lives on some open connected set $J \subset \mathbb{R}^m$, which plays the role that the interval (a, b) played in the one-dimensional case. The initial condition for the process is given by some $\mathbf{X}_0 \in J$. We denote again by T the first exit time from the set J

Calculating expectations of functions. For an arbitrary test function $f : \mathbb{R}^m \to \mathbb{R}$, we now define the function $F(\cdot) : \mathbb{R}^k \to \mathbb{R}$ analogously to before:

$$F(\mathbf{X}_0) = \mathbb{E}_0 \int_0^T f(\mathbf{X}_t)\mathrm{d}t.$$

Then, Ito's Lemma gives us the following approximation over a short $\mathrm{d}t$:

$$\begin{aligned} F(\mathbf{X}_t) =& f(\mathbf{X}_t)\mathrm{d}t + \mathbb{E}_t\Big[F(\mathbf{X}_{t+\mathrm{d}t})\Big] \\ =& f(\mathbf{X}_t)\mathrm{d}t + F(\mathbf{X}_t) + \mathbf{m}(\mathbf{X}_t)'\nabla F(\mathbf{X}_t) + \\ &+ \mathbf{\Sigma}(\mathbf{X}_t)\mathbb{E}_t\big[\mathrm{d}\mathbf{W}_t\big] + \mathbf{\Sigma}(\mathbf{X}_t)\nabla^2 F(\mathbf{X}_t)\mathbf{\Sigma}(\mathbf{X}_t)'\mathrm{d}t, \end{aligned}$$

where $'$ denotes the matrix transpose, ∇ denotes the gradient, and where ∇^2 denotes the Hessian. As in the scalar case, the expectation of all innovations $\mathbb{E}_t[\mathrm{d}W_t^{(i)}]$, $i \in \{1, \ldots, m\}$, is zero and we thus obtain the following PDE for $F(\cdot)$:

$$\mathbf{m}(\mathbf{X})'\nabla F(\mathbf{X}) + \mathbf{\Sigma}(\mathbf{X})\nabla^2 F(\mathbf{X})\mathbf{\Sigma}(\mathbf{X})' = -f(\mathbf{X}). \tag{34}$$

Example. An example is probably helpful to see what these PDEs look like in higher dimensions; we will also show here how to take care of discounting in this example. Let us imagine a firm whose state evolves according to the SDEs

$$\begin{aligned} \mathrm{d}z_t &= \rho z_t + \sigma \mathrm{d}W_t, \\ \mathrm{d}k_t &= I(z_t, k_t)\mathrm{d}t. \end{aligned}$$

Here, z_t is the firm's (log-)productivity that evolves according to an Ornstein-Uhlenbeck process with mean-reversion parameter $\rho \in (0, 1)$ and shock volatility $\sigma > 0$. k_t is capital of the firm and $I(\cdot)$ is a function that tells us the firm's investment given the state $\mathbf{X}_t \equiv (z_t, k_t)$. Let X be the state space; furthermore, let $A \subset X$ be the (closed) set of states at which the firm exits the market and let $B \subset X$ be the (closed) set of points at which the firm goes public. Let the set $J \equiv X \setminus (A \cup B)$ denote the open set on the firm continues operating.

Let us now find the PDE that expected profits of the firm have to obey. We define expected discounted profits as

$$\Pi(z_0, k_0) = \mathbb{E}_0 \int_0^T e^{-rt}\pi(z_t, k_t)\mathrm{d}t,$$

where $\pi(\cdot)$ is a given function, $r > 0$ is the discount rate, and where T is the first time at which the firm hits either set A or set B. Now, we can use Ito's Lemma on the profit function $\Pi(\cdot)$ at an arbitrary point $(z_t, k_t) \in J$ – just as we did above – to find that

$$\begin{aligned}\Pi(z_t, k_t) =& \pi(z_t, k_t)\mathrm{d}t + \mathbb{E}_t\Big[e^{-r\mathrm{d}t}\Pi(z_{t+\mathrm{d}t}, k_{t+\mathrm{d}t})\Big] \\ =& \pi(z_t, k_t)\mathrm{d}t + \Pi(z_t, k_t) - r\Pi(z_t, k_t)\mathrm{d}t + I(z_t, k_t)\Pi_k(z_t, k_t)\mathrm{d}t \\ &+ \rho z_t \Pi_z(z_t, k_t)\mathrm{d}t + \mathbb{E}_t\big[\Pi_z(z_t, k_t)\mathrm{d}W_t\big] + \tfrac{\sigma^2}{2}\Pi_{zz}(z_t, k_t)\mathrm{d}t.\end{aligned}$$

We see that the stochastic terms only show up for productivity since capital is not subject to shocks (at least not directly). From this, we find that the function $\Pi(\cdot)$ has to obey the following PDE for all $(z, k) \in J$:

$$r\Pi(z, k) = \pi(z, k) + I(z, k)\Pi_k(z, k) + \rho z \Pi_z(z, k)\mathrm{d}t + \tfrac{\sigma^2}{2}\Pi_{zz}(z, k),$$

which is like an HJB without a max-operator. The boundary condition for the PDE is given by $\Pi(z, k) = 0$ for all $(z, k) \in A \cup B$; recall that we said we would only count profits up to the exit time from the set J.

Expected exit time. As in the one-dimensional case, we can obtain the expected exit time, $S(\mathbf{X}_0)$, by setting $f(\mathbf{X}) = 0$ in the PDE (4).

Hitting probabilites for sets. In higher dimensions, the equivalent to the probability of a processes exiting an interval above instead ot below is the probability that a process hits a certain set $B \subset \mathbb{R}^n$ before it hits another set $A \subset \mathbb{R}^n$. In our firm example, let's define the probability $P(\mathbf{X}_0)$ that the firm goes public (i.e. it hits set A) before it goes broke (i.e. it hits set B). $P(\cdot)$ has to obey the following equation for any interior $\mathbf{X}_0 \in J$ and short $\mathrm{d}t$:

$$P(\mathbf{X}_t) = \mathbb{E}_t\Big[P(\mathbf{X}_{t+\mathrm{d}t})\Big].$$

Applying Ito's Lemma as before gives us the following PDE for $P(\cdot)$:

$$\mathbf{m}(\mathbf{X})'\nabla P(\mathbf{X}) + \mathbf{\Sigma}(\mathbf{X})\nabla^2 P(\mathbf{X})\mathbf{\Sigma}(\mathbf{X})' = 0,$$

where the boundary conditions are

$$\begin{aligned} P(\mathbf{X}) &= 1 \qquad && \text{for all } \mathbf{X} \in B, \\ P(\mathbf{X}) &= 0 && \text{for all } \mathbf{X} \in A. \end{aligned}$$

In the firm example, this PDE is simply

$$I(z, k)P_k(z, k) + \rho z P_z(z, k) + \tfrac{\sigma^2}{2}P_{zz}(z, k) = 0.$$

Green functions in higher dimensions. Green Functions can also be constructed in higher dimensions (google them and you will find some spike-shaped functions for the two-dimensional case). However, I haven't run into an economic application so far for which it is worth the effort to construct them;

usually it is easier to just solve the problem at hand numerically. For example, to find the expected exit time from a set or to calculate a discounted integral of some quantity, it is faster to solve the specific Kolmogorov Backward Equations given above than to construct Green Functions.

Densities. When it comes to computing densities, in practice it is very easy to map them forward in time on a discrete grid using the Markov-chain-approximation method. Just as in one dimension, the density $g(\mathbf{X}_t, t)$ has to fulfill a *Kolmogorov Forward Equation*. I don't state it here since this is painful to do for the general process given above and doesn't give us much insight, I find.

What *can* give you some interesting insights is the following: When you forward-iterate densities on a grid, write out the transition matrix of the discretized process as a matrix (just stack all grid points in one big vector to do this). The resulting matrix will be highly sparse since the discretized process can only jump to neighboring states over a time increment Δt. Make sure that you use routines that take advantage of this sparsity in order not to perform lots of unnecessary multiplications by zero.

Now, have a close look at the transition matrix $\mathbf{H}$ that you have obtained. Just as is the case for discrete Markov chains, the transpose $\mathbf{H}'$ is the linear operator that you would use in the Kolmogorov Backward Equation when calculating expectations of functions for $t + \Delta t$ for a given state $\mathbf{X}_t$ at t. This is the famous *adjoint property* of the two Kolmogorov equations; it says that the partial differential operator in the *forward* equation is the *adjoint* of the partial differential operator showing up in the *backward* equation. An *adjoint* is a generalization of a matrix transpose to function spaces.

References

Breiman, L. (1992), 'Probability, Volume 7 of Classics in Applied Mathematics', *Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA* .

Helland, I. (1996), 'One-dimensional diffusion processes and their boundaries', *Preprint series. Statistical Research Report http://urn. nb. no/URN: NBN: no-23420* .

Luttmer, E. G. (2007), 'Selection, growth, and the size distribution of firms', *The Quarterly Journal of Economics* **122**(3), 1103–1144.