Week 8
Stopping times, martingales, strategies

Jonathan Goodman

November 12, 2012

1 Introduction to the material for the week

Suppose X_t is a stochastic process and S is some set. The *hitting time* is the first time X_t hits S.

$$\tau = \min \{t \mid X_t \in S\} \ . \tag{1}$$

This definition makes sense without extra mathematical technicalities if X_t is a continuous function of t and S is a closed set.[1] In that case, $X_\tau \in S$ and $X_t \notin S$ if $t < \tau$. Many practical problems may be formulated using hitting times. When does something break? How long does it take to travel a given distance?

A hitting time is an important example of the more general idea of a *stopping time*. A stopping time is a time that depends on the path $X_{[0,T]}$, which makes it a random variable. What distinguishes a stopping time is that you know at time t whether $\tau \leq t$. If $\mathcal{F}_t$ is the filtration corresponding to X_t, then

$$\{\tau \leq t\} \in \mathcal{F}_t \ . \tag{2}$$

A hitting time is a stopping time because at time t you know all the values X_s for $s \leq t$, so you know whether $X_s \in S$ for some $s \leq t$. There are stopping times that are not hitting times. For example, you could stop the first time X_t has been inside S for a given amount of time.

Some random times are not stopping times. For example, take the maximum time rather than the minimum time in (1). This would be the "last exit time" for S. At time t, you may not know whether X_s will enter S for some $s > t$, so you do not know whether $\tau \leq t$.

Stopping times give a way to model *optimal decision problems* related to the stochastic process X_t. An optimal decision problem is the problem of finding the function $\tau(X_{[0,T]}$ that maximizes or minimizes some performance criterion. The early exercise problem for American style stock options is an optimal stopping

[1] The set S is *closed* if S includes all its limit points. If $x_n \in S$ and $x_n \to x$ as $n \to \infty$, then $x \in S$. For example, in one dimension, $S = \{0 < x < 1\}$ is not closed because $x_n = 1/n$ converges to $x = 0$, but $0 \notin S$.

problem. Many clinical drug trials have stopping criteria that end the trial if the drug quickly shows itself to be helpful, or dangerous.

Consider the problem of stopping a Brownian motion at the largest value possible

$$\max_{\tau(W_{[0,T]})} \mathrm{E}[\, W_\tau] \ . \tag{3}$$

One possible solution would be to take the actual maximum value:

$$W_\tau = \max_{0 \le t \le T} W_t \ .$$

But this is not a stopping time. Even if W_t is the largest value you have seen so far, you have no way of knowing whether $W_s > W_t$ for some $s > t$. (*Correction*: You do know. Almost surely there is an $s \in (t, T)$ with $W_s > W_t$.) The optimal decision problem would be to restrict the class of random times to those that are stopping times. You have to say at time t whether to stop at t or keep going.

Many hitting time problems and optimal decision problems may be solved using partial differential equations. For hitting time problems, you solve the forward or backward equation in the complement of S and specify a *boundary condition* at the boundary of S. Many optimal decision problems have the structure that the optimal stopping time is given by an optimal *decision boundary*. This is a set S_t so that τ is the first time $X_t \in S_t$.

A stochastic process is a *martingale* if, for any $s \ge t$,

$$\mathrm{E}[\, X_s \mid \mathcal{F}_t] = X_t \ . \tag{4}$$

If X_t is a diffusion process, then it is a martingale if the drift coefficient is equal to zero. That is

$$\mathrm{E}[\, dX_t \mid \mathcal{F}_t] = 0 \ .$$

A general theorem of Doob states that if f_t is an adapted process and if f and X are not too "wild", then

$$Y_t = \int_0^t f_s \, dX_s$$

is also a martingale. In some sense this is obvious, because the drift coefficient of Y is

$$\mathrm{E}[\, dY_t \mid \mathcal{F}_t] = f_t \mathrm{E}[\, dX_t \mid \mathcal{F}_t] = 0 \ .$$

if X is a martingale. The value f_t is known at time t if f_t is adapted. How wild is too wild? That's not a question for this course. But we give some examples where it is true and false.

The *Doob stopping time theorem* is a special case of the general martingale theorem. If τ is a stopping time that satisfies $\tau \le T$ (almost surely), then

$$\mathrm{E}[\, X_\tau] = x_0 \ . \tag{5}$$

To prove this, let Y_t be the *stopped* process: $Y_t = X_t$ if $t \le \tau$ and $Y_t = X_\tau$ for $t \ge \tau$ (the definitions agree for $t = \tau$). If $\tau \le T$, then $Y_T = X_\tau$. The trick is

to write Y as a stochastic integral involving X. The integrand is the switching function $f_t = 1$ for $t \leq \tau$ and $f_t = 0$ for $t > \tau$. This is an adapted function – you can determine the value of f_t knowing only $X_{[0,t]}$. If

$$Y_t = x_0 + \int_0^t f_s dX_s \ ,$$

then $Y_t = X_t$ if $t \leq \tau$ and Y_t is constant for $t > \tau$. The Doob martingale theorem implies that Y_t is a martingale. Therefore

$$\mathrm{E}[\, Y_T] = \mathrm{E}[\, X_\tau] = x_0 \ .$$

2 Backward equation boundary conditions

There are many questions involving hitting times for a set S that can be answered using a value function $f(x,t)$. The PDE for f depends on the process X, but not on S. The set S determines *boundary conditions* that f must satisfy. If you do it right, f will be completely determined by the final condition, the boundary conditions, and the PDE.

The PDE involves the *generator* of the process X_t, which is a *differential operator*. For a given t, think of $f(\cdot, t)$ as an abstract *vector*. For a one dimensional diffusion, L acts on the vector f as

$$Lf(x,t) = \frac{1}{2}\mu(x,t)\partial_x^2 f(x,t) + a(x,t)\partial_x f(x,t) \ . \tag{6}$$

The generator L does not act on the t variable, so t is just a parameter that says which function $f(\cdot, t)$ the generator is acting on.[2]

For example, if X is an Ornstein Uhlenbeck process with parameters σ^2 and γ, then $\mu = \sigma^2$ and $a = -\gamma x$, so

$$Lf = \frac{1}{2}\sigma^2 \partial_x^2 f - \gamma x \partial_x f \ .$$

If $f(x,t) = 3x^2 - tx + t^2$, then $Lf = 3\sigma^2 - 6\gamma x^2 + \gamma t x$. The operator L may be expressed as

$$L = \frac{1}{2}\mu(x,t)\partial_x^2 + x(x,t)\partial_x \ . \tag{7}$$

Then applying L to a function f is given by the expression (6). Mathematicians say that the operator L "sends f to Lf", or that f "goes to" Lf:

$$f \xrightarrow{L} Lf = \frac{1}{2}\sigma^2 \partial_x^2 f - \gamma x \partial_x f \ .$$

For example, one might write

$$e^{-x^2/2} \xrightarrow{\partial_x^2} \left(x^2 - 1\right) e^{-x^2/2} \ .$$

[2] A famous joke defines a *parameter* as a variable constant.

Suppose the diffusion is multi-dimensional. Let n be the number of components of X_t. The generator in this case is

$$Lf = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_{ij}(x,t)\partial_{x_i}\partial_{x_j}f + \sum_{i=1}^{n}a_i(x,t)\partial_{x_i}f \ . \tag{8}$$

We will do some multi-dimensional examples at some point.

Here is a simple example. Let X_t be a one dimensional diffusion with $x_0 = 1$. Let $V(x)$ be a payout function, and suppose you get payout $V(X_T)$ only if $X_t > 0$ for $0 \leq t \leq T$. We need some notation for the mathematical formulation. The hitting time is $\tau = \min\{t \mid X_0 = 0\}$. The event we need to describe is the event that $\tau \geq T$. The *indicator function* of this event is $\mathbf{1}_{\tau\geq T}(X_{[0,T]})$, which has the values

$$\mathbf{1}_{\tau\geq T}(X_{[0,T]}) = \begin{cases} 1 & \text{if } \tau \geq T \\ 0 & \text{if } \tau < T. \end{cases}$$

(This is also called *characteristic function* and written $\chi_{\tau\geq T}$. We use the term "indicator function" because in probability, *characteristic function* can refer to Fourier transform.) In this notation, the payout $V(X_T)\,\mathbf{1}_{\tau\geq T}(X_{[0,T]})$. This is a function that is equal to $V(X_T)$ if $\tau \geq T$ and is equal to zero otherwise. The expected payout is

$$\mathrm{E}\big[\,V(X_T)\,\mathbf{1}_{\tau\geq T}(X_{[0,T]})\big] \ . \tag{9}$$

The PDE approach to calculating the expected payout (9) is to define a value function that satisfies a backward equation with boundary conditions. The value function that works is

$$f(x,t) = \mathrm{E}_{x,t}\big[\,V(X_T)\,\mathbf{1}_{\tau\geq T}(X_{[0,T]}) \mid \tau \geq t\big] \ . \tag{10}$$

If we can evaluate the value function f, we can plug in $x = 1$ and $t = 0$ to get a formula for (9):

$$f(1,0) = \mathrm{E}\big[\,V(X_T)\,\mathbf{1}_{\tau\geq T}(X_{[0,T]})\big] \ .$$

We compute the entire value function (10) for the purpose of getting the single number (9).

The value function defined in (10) may seem more complicated than necessary. A function that is simpler to write down is

$$g(x,t) = \mathrm{E}_{x,t}\big[\,V(X_T)\,\mathbf{1}_{\tau\geq T}(X_{[0,T]})\big] \ . \tag{11}$$

The difference between these is that g counts paths that have touched zero at some time $s < t$. The definition (10) excludes such paths. More precisely, it conditions on not having them. The two definitions are related by Bayes' rule. In the the case here, the integrand is zero if $\tau < T$, so

$$f(x,t) = \frac{g(x,t)}{\mathrm{P}_{x,t}(\tau \geq T)} \ .$$

The definition of g is suitable for expressing as a conditional expectation, conditional on $\mathcal{F}_t$:

$$g(X_t, t) = \mathrm{E}\big[\, V(X_T)\, \mathbf{1}_{\tau \geq T}(X_{[0,T]}) \mid \mathcal{F}_t \big]$$

The denominator has a similar expression. In fact

$$\mathrm{P}(\tau \geq t) = \mathrm{E}\big[\, \mathbf{1}_{\tau \geq t}(X_{[0,T]}) \big] \ ,$$

so

$$f(X_t, t) = \frac{\mathrm{E}\big[\, V(X_T)\, \mathbf{1}_{\tau \geq T}(X_{[0,T]}) \mid \mathcal{F}_t \big]}{\mathrm{E}\big[\, \mathbf{1}_{\tau \geq t}(X_{[0,T]}) \mid \mathcal{F}_t \big]} \ .$$

This definition might be very hard to work with for the following reason. The denominator $\mathrm{E}\big[\, \mathbf{1}_{\tau \geq t}(X_{[0,T]}) \mid \mathcal{F}_t \big]$ is not really an expectation because the random variable $\mathbf{1}_{\tau \geq t}(X_{[0,T]})$ is known at time t. Therefore

$$\mathrm{E}\big[\, \mathbf{1}_{\tau \geq t}(X_{[0,T]}) \mid \mathcal{F}_t \big] = \mathbf{1}_{\tau \geq t}(X_{[0,T]}) \ .$$

But the formula

$$f(X_t, t) = \frac{\mathrm{E}\big[\, V(X_T)\, \mathbf{1}_{\tau \geq T}(X_{[0,T]}) \mid \mathcal{F}_t \big]}{\mathbf{1}_{\tau \geq t}(X_{[0,T]})}$$

looks bad, because the denominator is not a function of X_t and t alone. It depends on the path before t. Somehow, when you did the division, this dependence cancels out. The bottom line, for me, is that the more complicated definition (10) will be easier to work with.

The value function is found by solving a PDE problem. A PDE problem consists of a PDE and other conditions as appropriate – initial conditions, boundary conditions, final conditions, etc. The PDE in this case is the backward equation

$$\partial_t f = Lf = \frac{1}{2}\mu(x,t)\partial_x^2 f(x,t) + a(x,t)\partial_x f(x,t) \ . \tag{12}$$

This PDE is satisfied in the region $x > 0$. The value function may not be defined for $x < 0$. If it is defined, the most natural definition would be $f = 0$. Either way, the PDE (12) is used only in the *continuation region* $x > 0$. The final condition is clear from the definition of f. If $t = T$, and $x \geq 0$, then $f(x,T) = V(x)$. There is an extra boundary condition at $x = 0$, which is the boundary of the continuation region in this example. We can guess this boundary condition by continuity. If x is actually on the boundary of the continuation region, which is to say $x = 0$, then the definition (10) gives the value zero. If $f(x,t)$ is a continuous function of x, then 0 is the limiting value of f as $x \to 0$. This suggests that the boundary condition should be

$$f(0,t) = 0 \ . \tag{13}$$

Here is one of the few examples where f may be calculated explicitly. Let the process X_t be Brownian motion starting at $x_0 = 1$ but having $\mathrm{var}(X_t) = t$.

This makes $\sigma^2 = \mu$ 1, and $a = 0$. Take $V(x) = 1$, so f is just the conditional probability of not touching the boundary before time T. The PDE problem is: Find $f(x,t)$, defined for $x \geq 0$ and $t \leq T$ that satisfies the PDE

$$\partial_t f + \frac{1}{2}\partial_x^2 f = 0$$

where it is defined. In addition, f should satisfy the final condition $f(x,T) = 1$ for $x \geq 0$, and the boundary condition $f(0,t) = 0$ for $t \leq T$.

This problem may be solved using something like the method of images. We *extend* the definition of f so that f is defined for all x with the anti-symmetry condition $f(-x,t) = -f(x,t)$. If f is continuous, this implies that $f(0,t) = 0$. In order to achieve the skew-symmetry condition, we take the final condition to be skew symmetric. We do this without changing the already known values of $f(x,T)$ for $x > 0$. Clearly, the extended final condition should be $f(x,T) = 1$ for $x > 0$ and $f(x,t) = -1$ for $x < 0$. The value of f when $x = 0$ is irrelevant. There is no boundary. We are talking about the simple heat equation (OK, with the direction of time reversed). The solution may be given as a Green's function integral using the known final values:

$$\begin{aligned}
f(x,t) &= \int_{-\infty}^{\infty} G(x-y, T-t) f(y,T)\, dy \\
&= \int_0^{\infty} G(x-y, T-t)\, dy - \int_{-\infty}^{0} G(x-y, T-t)\, dy \\
&= \int_0^{\infty} \frac{1}{\sqrt{2\pi(T-t)}} e^{-(x-y)^2/2(T-t)}\, dy \\
&\quad - \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi(T-t)}} e^{-(x-y)^2/2(T-t)}\, dy\ .
\end{aligned}$$

The two Gaussian integrals on the last line represent probabilities. We can express them in terms of the cumulative normal distribution function $N(z) = \mathrm{P}(Z \leq z)$, where $Z \sim \mathcal{N}(0,1)$. The second integral on the last line is the probability that the random variable $Y \sim \mathcal{N}(x, T-t)$ has $Y < 0$. In general, if $Y \sim \mathcal{N}(\mu, \sigma^2)$, then $Y \sim \mu + \sigma Z$, where $Z \sim \mathcal{N}(0,1)$. The expression $Y \sim \mu + \sigma Z$ means that Y and $\mu + \sigma Z$ have the same distribution. This implies that

$$\mathrm{P}(Y < 0) = \mathrm{P}(\mu + \sigma Z < 0) = \mathrm{P}\left(Z < \frac{-\mu}{\sigma}\right) = N\left(\frac{-\mu}{\sigma}\right)\ .$$

In this example, $\mu = x$ and $\sigma = \sqrt{T-t}$, so

$$\mathrm{P}(Y < 0) = N\left(\frac{-x}{\sqrt{T-t}}\right)\ .$$

There are two properties of Gaussians, each of which would give a way to write the first integral in terms of N. The first is $\mathrm{P}(Z > a) = \mathrm{P}(-Z < -a) =$

$\mathrm{P}(Z < a)$, the last is because $-Z \sim Z$ – the Gaussian distribution is symmetric. This gives $\mathrm{P}(Z > a) = N(-a)$. In the present example, the first integral is

$$\mathrm{P}(Y > 0) = \mathrm{P}(\mu + \sigma Z > 0) = \mathrm{P}\left(Z > \frac{-\mu}{\sigma}\right) = N\left(\frac{\mu}{\sigma}\right) = N\left(\frac{x}{\sqrt{T-t}}\right) \ .$$

The resulting formula for the *survival probability* is

$$f(x,t) = \mathrm{P}_{x,t}(\tau > T) = N\left(\frac{x}{\sqrt{T-t}}\right) - N\left(\frac{-x}{\sqrt{T-t}}\right) \ . \tag{14}$$

Here is a quick check that this function satisfies all the conditions we set for it. It satisfies the PDE (a calculation using $N'(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$). It satisfies the boundary condition. If you put $x = 0$, the two terms on the right cancel exactly. It satisfies the final condition. If $x > 0$ and you send t to T, then $N\left(\frac{x}{\sqrt{T-t}}\right) \to 1$ and $N\left(\frac{-x}{\sqrt{T-t}}\right) \to 0$. That is because $\frac{x}{\sqrt{T-t}} \to \infty$ and $\frac{-x}{\sqrt{T-t}} \to -\infty$ (this is where you use $x > 0$).

The other fact about N is $\mathrm{P}(Z > a) = 1 - \mathrm{P}(Z < a) = 1 - N(a)$. Therefore,

$$\mathrm{P}\left(Z > \frac{-\mu}{\sigma}\right) = 1 - \mathrm{P}\left(Z < \frac{-\mu}{\sigma}\right) = 1 - N\left(\frac{-x}{\sqrt{T-t}}\right)$$

This gives a formula equivalent to (14), which is

$$f(x,t) = \mathrm{P}_{x,t}(\tau > T) = 1 - 2N\left(\frac{-x}{\sqrt{T-t}}\right) \ . \tag{15}$$

This formula is what you would get from the Kolmogorov reflection principle:

$$\mathrm{P}_{x,t}(X_s < 0 \text{ for some } s \in [t,T]) = 2\,\mathrm{P}_{x,t}(X_T < 0) \ ,$$

so

$$\begin{aligned} \mathrm{P}_{x,t}(X_s > 0 \text{ for all } s \in [t,T]) &= 1 - \mathrm{P}_{x,t}(X_s < 0 \text{ for some } s \in [t,T]) \\ &= 1 - 2\,\mathrm{P}_{x,t}(X_T < 0) \\ &= 1 - 2\,N\left(\frac{-x}{\sqrt{T-t}}\right) \ . \end{aligned}$$

The formula (15) satisfies the PDE for the same reason as (14). It satisfies the final condition because $\mathrm{P}(X_T < 0 \mid X_T = x > 0) = 0$. It satisfies the boundary condition because $N(0) = \frac{1}{2}$.

We can use (14) or (15) to estimate the survival probability starting from a fixed x at time $t = 0$ as $T \to \infty$. This is the probability of not hitting the boundary for a long time. The argument to N goes to zero as $T \to \infty$. Therefore, we use $N(\epsilon) \approx N(0) + N'(0)\epsilon$. We already saw that $N(0) = \frac{1}{2}$ and $N'(0) = \frac{1}{\sqrt{2\pi}}$. Therefore, for large T,

$$\mathrm{P}_{x,0}(\tau > T) \approx \frac{x}{\sqrt{2\pi T}} \ . \tag{16}$$

We see that this goes to zero as $T \to \infty$. Therefore, we know that from any starting point, Brownian motion hits $x = 0$ at some positive time almost surely. This is true also in two dimensions – a two dimensional Brownian motion will touch the origin almost surely. In three or more dimensions it is not true. In fact, if $|X_0| > 1$, there is a positive probability that $|X_t| > 1$ for all $t > 0$. Brownian motion in one or two dimensions is *recurrent*, while it is *transient* in dimensions 3 or more.

While we are talking about these solutions to the backward equation, let us notice some other properties. One is the *smoothing* property. The formula (14) defines a function that is discontinuous when $t = T$. Nevertheless, $f(x, t)$ is a smooth function of x for $t < T$. This is a general property of PDE's of diffusion type.

Some other properties are illustrated by a different solution of the backward equation

$$h(x,t) = N\left(\frac{x+1}{\sqrt{T-t}}\right) - N\left(\frac{x-1}{\sqrt{T-t}}\right) .$$

This function has final values $h(x, T) = 1$ if $-1 < x < 1$ and $h(x, T) = 0$ otherwise. That makes $h(x, T) = \mathbf{1}_{|x|<1}$, which is a *step function* that is different from zero when x is not too far from zero. Whenever $t < T$, $h(x, t) > 0$ for any x. This means that the fact that $h > 0$ "propagates" infinitely fast through the whole domain where h is defined. This is also a property of general diffusion PDE's. However, the solution is not large for $x > 1$ and t close to T. In fact, it is exponentially small. The influence is very small in short times and large distances.

Finally, look at $h(x, t)$ for x near -1 and t close to T. The second term is exponentially small, as we just said. But the first term looks like the solution with final data that have a jump at $x = -1$. The behavior near -1 and T is almost completely determined by the final condition there. This is approximate locality.