1 Continuous-Time Stochastic Processes

We are going to begin looking at models that are developed in continuous time now. All of these models are built with what is called continuous-time stochastic processes, or stochastic calculus. The purpose of this section of the notes is to introduce us to continuous-time math. Some aspects of continuous-time math are a little bit counter-intuitive, so be prepared for some strange stuff. Good references on continuous-time math are the books by Shimko and Merton, both of which are listed in the syllabus. Another, more advanced reference is *Brownian Motion and Stochastic Calculus* by Ioannis Karatzas and Steven Shreve (published by Springer-Verlag, New York). Most of the material in this note is from Shimko's book.

1.1 Stochastic Processes

The analysis of stochastic processes is a topic in mathematical statistics. A stochastic process amounts to a series of data, usually a time-series. We typically make distributional assumption about the series we are interested in analyzing. We then determine what the implications of our distributional assumption are. If our series is $X_0, X_1, X_2, \ldots, X_T$ then we can express the distribution of X_t conditional on the first $t-1$ values of X as $f_{t-1}(X_t|x_0, x_1, \ldots, x_{t-1})$. Our stochastic process becomes a continuous-time process as the distance (or the time) between period t (when X_t is observed) and period $t-1$ becomes arbitrarily small.

A common example of a discrete-time stochastic process is the random walk,

$$X_t = \mu + X_{t-1} + \varepsilon_t, \tag{1}$$

where the ε_t are iid. The parameter μ, which is the expected change per period, is called the *drift* term. Another example of a stochastic process is the autoregressive process,

$$X_t = (1-a)\mu + aX_{t-1} + \varepsilon_t, \tag{2}$$

where the errors are again iid. As mentioned previously, a Markov process is a process in which all the relevant information about X_t is in x_{t-1},

$$f_{t-1}(X_t|x_0, x_1, \ldots, x_{t-1}) = f_{t-1}(X_t|x_{t-1}) \tag{3}$$

Both the random walk (1) and the autoregressive process (2) are examples of Markov processes.

A *martingale* is a stochastic process with the properties

$$\begin{aligned} &E[|X_t|] < \infty \ \forall t \\ &E[X_t|x_0, x_1, \ldots, x_{t-1}] = x_{t-1} \end{aligned} \tag{4}$$

The random walk (1) with zero drift, or $\mu = 0$, is an example of a martingale. The autoregressive process (2) is also a martingale if $a = 1$. More generally, the stochastic process $\{X_t\}_0^T$ is a martingale with respect to the information in the stochastic process $\{Y_t\}_0^T$ if, in addition to (4),

$$E[X_t|y_0, y_1, \ldots, y_{t-1}] = x_{t-1}. \tag{5}$$

A time series of updated conditional expectations is always a martingale.

1.2 Standard Wiener Process

Now let's think about allowing the time between observations to become arbitrarily small. In particular, let's think about allowing a special kind of random walk to occur in continuous time. Consider a random walk that has normally distributed errors and no drift term,

$$W(t+\Delta) = W(t) + \varepsilon(t+\Delta), \quad W(0) = w_0, \quad \varepsilon(t+\Delta) \sim \text{ iid. } \mathcal{N}(0, \Delta). \tag{6}$$

Notice that the variance of the error terms depends on the time period over which the observation is taken. Taking the limit of the difference of this process as Δ goes to zero produces what is known as the standard Weiner process.

$$dW(t) = \lim_{dt \to 0} W(t+dt) - W(t) = \lim_{dt \to 0} \varepsilon(t+dt), \quad \varepsilon(t+dt) \sim \text{ iid } \mathcal{N}(0, dt). \tag{7}$$

We will heuristically define dt as the smallest positive real number such that $dt^\alpha = 0$ whenever $\alpha > 1$. This process is also sometimes called white noise. The Wiener process is the basic building block of stochastic calculus, and it has some really strange properties. We need to understand it somewhat thoroughly.

Some of the properties of the Wiener process are listed here and discussed below:

1. $E[dW(t)] = 0$
2. $E[dW(t)dt] = E[dW(t)]dt = 0$
3. $E[dW(t)^2] = dt$
4. $\text{VAR}[dW(t)^2] = E[dW(t)^4] - E^2[dW(t)^2] = 3dt^2 - dt^2 = 0$
5. $E[(dW(t)dt)^2] = E[dW(t)^2]dt^2 = 0$
6. $\text{VAR}[dW(t)dt] = E[(dW(t)dt)^2] - E^2[dW(t)dt] = 0$
7. $dW(t)^2 = dt$
8. $dW(t)dt = 0$
9. $W(t)$ is continuous in t

10. $W(t)$ is nowhere differentiable

11. $W(t)$ is a process of unbounded variation

12. $W(t)$ is a process of bounded quadratic variation

13. The conditional distribution of $W(u)$ given $W(t)$ for $u > t$ is normal with mean $W(t)$ and variance $(u - t)$

14. The variance of a forecast of $W(u)$ increases indefinitely as $u \to \infty$

We should discuss each of these properties in turn. To understand these properties, it is important to keep in mind that dt is deterministic, while $dW(t)$ is a normally distributed random variable. Properties (1) though (4) follow from the distribution of $dW(t)$. Property (5) follows from the fact that dt is such a small number that raising it to any power greater than one makes it disappear ($dt^2 = 0$). Property (6) follows from property (5) and property (2). Since the variances of $dW(t)^2$ and $dW(t)dt$ are both zero, we can say that these two terms are always equal to their expectations, as properties (7) and (8) state. Properties (9) through (14) describe the distribution of $W(t)$ in a little more detail.

Property 7 is the basis for a lot of people's intuition about Wiener processes. People often like to think of dW as the square root of dt. This is not strictly correct, of course, but it does make for a simple rule of thumb for determining which elements of a stochastic process are important. If you think of dW as $\sqrt{dt}$ then you simply need to apply the rule that dt raised to any power greater than 1 is zero to get the right answer most of the time.

It should be noted that we are only discussing one particular representation of a Wiener process here. Continuous-time stochastic processes can either be expressed in differential form, as we have expressed the Wiener process here, or they can be expressed in stochastic integral form,

$$W(t) = w_0 + \int_0^t dW(u). \tag{8}$$

The stochastic integral form is preferred by people who prove theorems about stochastic calculus. We will use the differential form almost exclusively - it is a little more "engineer" (less rigorous) but it will suffice for us.

The standard Wiener process is not sufficiently general for us to use. We will frequently use the generalized Wiener process,

$$dX = \alpha(X, t)dt + \sigma(X, t)dW \tag{9}$$

where α is now a drift function and σ is a volatility function that can depend on the level of X and time, t. By choosing different functions for the drift and volatility, we can generate a lot of different stochastic processes.

1.3 Common Continuous-Time Processes

Three common stochastic processes are discussed in this subsection.

1.3.1 Arithmetic Brownian Motion

$$dX = \alpha dt + \sigma dW \tag{10}$$

- grows at a linear rate
- X may be positive or negative
- the distribution of $X(u)$ given $X(t)$ is $\mathcal{N}(X(t) + \alpha(u - t), \sigma^2(u - t))$
- the variance of a forecast of $X(u)$ goes to infinity as $u \to \infty$.

- arithmetic Brownian motion is often used to model continuously-compounded or logarithmic returns

1.3.2 Geometric Brownian Motion

$$dX = \alpha X dt + \sigma X dW \tag{11}$$

- grows at an exponential rate
- volatility is proportional to level of X
- if X starts out positive, it will remain positive
- X has an absorbing barrier at 0 - if it hits 0 (a zero probability event) it will stay there
- the distribution of $X(u)$ given $X(t)$ is lognormal

  the mean of $\ln X(u)$ is $\ln X(t) + \alpha(u-t) - \frac{1}{2}\sigma^2(u-t)$

  the standard deviation of $\ln X(u)$ is $\sigma\sqrt{(u-t)}$
- $\ln X(u)$ is normally distributed

  the expected value of $X(u)$ conditional on $X(t)$ is $X(t)e^{\alpha(u-t)}$
- the variance of a forecast of $X(u)$ goes to infinity as $u \to \infty$.
- geometric Brownian motion is often used to model security prices

1.3.3 Mean Reverting Process

$$dX = \kappa(\mu - X)dt + \sigma X^{\gamma} dW \tag{12}$$

- also called the Ornstein-Uhlenbeck process when $\gamma = 1$.
- constantly reverting to long run mean, μ

- reverts to μ at a speed of $\kappa > 0$
- volatility depends on both $\sigma > 0$ and $\gamma > 0$
- if X starts out positive, it will remain positive
- as X approaches 0 it has a positive drift and no volatility
- if $\gamma = \frac{1}{2}$ then the distribution of $X(u)$ given $X(t)$ is non-central χ^2

  the mean is $(X(t) - \mu)e^{-\kappa(u-t)} + \mu$

  the variance is $\frac{X(t)\sigma^2}{\kappa}[e^{-\kappa(u-t)} - e^{-2\kappa(u-t)}] + \frac{\mu\sigma^2}{2\kappa}[1 - e^{-\kappa(u-t)}]^2$
- the variance of a forecast of $X(u)$ is finite as $u \to \infty$.
- the mean reverting process is often used to model interest rates

1.4 Ito's Lemma

Now let's think about functions of Wiener processes. Using a Taylor series expansion, we can express $f(X + \Delta)$ as

$$f(X + \Delta) = f(X) + \Delta f_X(X) + \frac{1}{2}\Delta^2 f_{XX}(X) + \frac{1}{6}\Delta^3 f_{XXX}(X) + \dots \tag{13}$$

Standard calculus lets Δ approach dX and thus lets the second order and higher terms vanish ($dX^2 \to 0$). In standard calculus,

$$f(X + dX) = f(X) + f_X(X)dX \quad \text{or} \quad df(X) = f_X(X)dX. \tag{14}$$

We need to think carefully about what happens to higher order terms in stochastic calculus. Let's think about the term involving dX^3:

$$
\begin{aligned}
dX^3 &= dX[\alpha(X,t)dt + \sigma(X,t)dW]^2 \\
&= dX[\alpha(X,t)^2 dt^2 + 2\sigma\alpha dW dt + \sigma(X,t)^2 dW^2] \\
&= dX\sigma(X,t)^2 dt \\
&= [\alpha(X,t)dt + \sigma(X,t)dW]\sigma^2 dt = 0.
\end{aligned} \tag{15}
$$

Clearly, the manner in which we determined that dX^3 is zero applies to dX raised to all powers greater than 3. Now let's think about the term with dX^2:

$$
\begin{aligned}
dX^2 &= [\alpha(X,t)dt + \sigma(X,t)dW]^2 \\
&= \alpha(X,t)^2 dt^2 + 2\sigma\alpha dW dt + \sigma(X,t)^2 dW^2 \\
&= \sigma(X,t)^2 dt
\end{aligned} \tag{16}
$$

In stochastic calculus, the dX^2 term does not vanish - it becomes a function of the deterministic quantity dt. This occurs because the dW^2 term in dX^2 is the square of a normally distributed random variable. Higher order terms do vanish because they are of order higher than dt. In stochastic calculus, the analog to (14) is

$$
df(X) = f_X(X)dX + \frac{1}{2}f_{XX}(X)dX^2. \tag{17}
$$

This equation is a simple statement of the famous Ito's lemma. We will use it many times in continuous-time model building. It can be extended in various ways. For example, if the function we are considering also happens to be a function of time, we can write

$$
\begin{aligned}
f &= f(X,t) \\
df &= f_X dX + f_t dt + \tfrac{1}{2}[f_{XX}dX^2 + 2f_{Xt}dXdt + f_{tt}dt^2] \\
&= f_X dX + f_t dt + \tfrac{1}{2}f_{XX}dX^2 \\
&= f_X dX + f_t dt + \tfrac{1}{2}f_{XX}\sigma(X,t)^2 dt \\
&= [\alpha(X,t)f_X + \tfrac{1}{2}\sigma(X,t)^2 f_{XX} + f_t]dt + \sigma f_X dW.
\end{aligned} \tag{18}
$$

We used the properties of the generalized Wiener process several times in this derivation.

What if we have more than one variable being generated by a generalized Wiener process? Then we use the multivariate form of Ito's lemma. We now have a system of processes that we can define as follows:

$$\begin{aligned} dX &= \alpha(X,Y,t)dt + \sigma(X,Y,t)dW \\ dY &= \beta(X,Y,t)dt + \nu(X,Y,t)dZ \\ dZdW &= \rho dt. \end{aligned} \tag{19}$$

Now we can think of a form of Ito's lemma for functions of both X and Y:

$$\begin{aligned} f &= f(X,Y,t) \\ df &= f_X dX + f_Y dY + f_t dt + \tfrac{1}{2}[f_{XX}dX^2 + 2f_{XY}dXdY + f_{YY}dY^2] \end{aligned} \tag{20}$$

Solving out the terms in brackets yields

$$df = [\alpha f_X + \beta f_Y + f_t + \frac{1}{2}\sigma^2 f_{XX} + \rho\sigma\nu f_{XY} + \frac{1}{2}\nu^2 f_{YY}]dt + \sigma f_X dW + \nu f_Y dZ \tag{21}$$

1.5 Poisson (Jump) Processes

The generalized Wiener process that we have discussed so far is fairly general, but it requires changes in X to occur continuously. There are a lot of economic variables that appear to jump rather than evolve continuously, so we need to talk about jump processes or Poisson processes. A Poisson process increases by one every time it jumps. We assume that it jumps with probability λ in a time period of length dt. We also assume that dt is a sufficiently short amount of time to preclude more than one jump.

We call changes in the Poisson process dq:

$$\begin{aligned} dq(t) &= 1 \text{ with probability } \lambda dt \\ &= 0 \text{ with probability } 1 - \lambda dt \end{aligned} \tag{22}$$

It is simplest to think of λ as a constant, but you can easily imagine allowing it to vary with X and t. In order to allow jumps of different sizes, we multiply our Poisson process term by a random variable with compact support (with a closed and bounded domain of positive probability) called $\zeta(X, t)$. Now we can write a process that has both continuous and discrete components as

$$dX = \alpha dt + \sigma dW + \zeta dq \tag{23}$$

ζ can take values like $-X$ to indicate possible default. Jump risk is often considered diversifiable is asset pricing models. This allows people to just worry about the expected value of a jump rather than assigning any risk premium to jump risk.

The appropriate version of Ito's lemma is now

$$df = f_X dX + f_t dt + 1/2 f_{XX} dX^2 + [f(X + \zeta) - f(X)]dq \tag{24}$$

The expected change in our function's value can be computed as

$$E[df] = \left(f_X \alpha + \frac{1}{2} f_{XX} \sigma^2 + f_t + \lambda E[f(X + \zeta) - f(X)] \right) dt \tag{25}$$

1.6 Homework Problems

1. Verify equation (21) by deriving it from (20).

2. Do problem 1 in Shimko's book.

3. Do problem 2 in Shimko's book.

4. Do problem 4 in Shimko's book.

5. Do problem 5 in Shimko's book.

6. Do problem 6 in Shimko's book.

7. Do problem 7 in Shimko's book.

You may want to read through the examples in Shimko's book before starting the problems.