Up to now, all general diffusion processes have been defined as functions of Brownian motion. The Ito integral produces X_t as a function of $W_{[0,t]}$ (our notation for the Brownian motion path values on the interval $[0,t]$). More precisely, we used the filtration $\mathcal{F}_t$ generated by $W_{0,t]}$ and created stochastic processes X_t that were measurable in $\mathcal{F}_t$. We called such processes non-anticipating, or adapted, or whatever. This meant that the value of X_t was determined by the values W_s for $0 \leq s \leq t$. We now turn to a more general diffusion process that is discussed on its own terms rather than in terms of a Brownian motion that might generate it.

A related idea is the difference between *strong* and *weak* solutions of stochastic differential equations, SDE's. An SDE takes the form

$$dX_t \;=\; a(X_t)dt \,+\, b(X_t)dW_t \;. \tag{1}$$

Models of continuous time dynamics with randomness often take this form. A strong solution is a progressively measurable function $X_t(W_{[0,t]}, t)$ so that

$$X_t \,-\, X_0 \;=\; \int_0^t a(X_s)\, ds \,+\, \int_0^t b(X_s)\, dW_s \;. \tag{2}$$

The favorite example is *geometric Brownian motion*, which is the SDE

$$dX_t \;=\; \mu X_t dt \,+\, \sigma X_t dW_t \;. \tag{3}$$

The solution is

$$X_t = X_0 e^{\sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right)t} \tag{4}$$

It is a simple exercise to take $u(w,t) = X_0 e^{\sigma w + \left(\mu - \frac{\sigma^2}{2}\right)t}$, compute the partial derivatives in Ito's lemma and check that they fit together in a way that (4) satisfies (3). This is the strong solution to (3), which is the solution given as a function of Brownian motion.

Before getting to weak solution, here is another piece of mathematical shorthand, the *little o* notation. We say that $f(t) = o(g(t))$ if $g \to 0$ as $t \to 0$, $g(t) > 0$ if $t > 0$, and $f(t)/g(t) \to 0$ as $t \to 0$. This just says that f is small relative to g as $t \to 0$. For example, if $f(0) = 0$ and f is differentiable at $t = 0$, then $f(t) = f'(0)\, t \,+\, o(t)$. This says that $\lim_{t\to 0} |f(t) - t\, f'(0)| \,/t = 0$. In other words, the error in the first derivative approximation is smaller than the approximation, or the error as a percent of the answer goes to zero with t. We use the little o notation to avoid making highly technical and ultimately irrelevant precise statements about the accuracy of an approximation. For example, if f has two derivatives, then $f(t) = f(0) + tf'(0) + O(t^2)$. This gives the sharper bound Ct^2 instead of just $o(t)$. In our applications it often is easy to get error bounds $O(t^{3/2})$, which is $o(t)$ but less so than $O(t^2)$ (whatever "less so" might mean). Instead of arguing about just how small the error term might be, I say $o(t)$ which is (a) true, (b) easy to verify, (c) enough of an error bound for our purposes.

A weak solution is just a stochastic process (probability space Ω, filtration $\mathcal{F}_t$, functions X_t measurable with respect to $\mathcal{F}_t$) that satisfies (1) in the sense that (use the notation $\Delta X = X_{t+\Delta t} - X_t$)

$$E[\,\Delta X \mid \mathcal{F}_t] \;=\; a(X_t)\Delta t \,+\, o(\Delta t)\;, \tag{5}$$

and

$$E\big[\,\Delta X^2 \mid \mathcal{F}_t\big] \;=\; b(X_t^2)\Delta t \,+\, o(\Delta t)\;, \tag{6}$$

and

$$E\big[\,\Delta X^4 \mid \mathcal{F}_t\big] \;=\; o(\Delta t)\;. \tag{7}$$

We verified last week that the strong solution that satisfies (2) satisfies (5) and (6). As it says above, the error bounds $o(\Delta t)$ are not *sharp*. The statements are true with $o(\Delta t)$ replaced by the stronger error bounds $O(\Delta t^2)$. The reason not to bother with that is that the stronger bounds are harder to prove, and the difference does not matter for the applications we have in mind.

The point of the weak formulation, rather than the strong Brownian motion formulation, is that one can determine the coefficients $a(x)$ and $b(x)$ directly from the application using (5) and (6). Here is an example of an SDE model derived using such reasoning. It is a queue with abandonment. At time t there are N_t *customers* in the queue. If $N_t > 0$, the *server* is at work on one of the customers. If the server is working, there is a probability μdt that the service finished in time interval dt. Another way to say this is to say that the service time is an exponential random variable with rate constant μ. In the language of continuous time Markov chains, this translates to

$$\Pr(N \to N-1 \text{ in time } dt \text{ because of service }) \;=\; \mu dt\;.$$

New customers arrive as a poisson process with rate constant λ. This means that the probability of a new arrival in time dt is λdt:

$$\Pr(N \to N+1 \text{ in time } dt \text{ because of arrival }) \;=\; \lambda dt\;.$$

Abandonment is when a waiting customer leaves without service. We model this with an abandonment rate r. A waiting customer abandons in time interval dt with probability rdt, with all abandonment decisions being independent. The result is

$$\Pr(N \to N-1 \text{ in time } dt \text{ because of abandonment }) \;=\; rN_t dt\;.$$

Given that all these events are unlikely in a small interval of time, it is even more unlikely that more than one event will happen. Therefore, we use only single event probabilities in the calculations below.

If N_t is large it is natural to make a model in which N_t is a continuous variable rather than discrete. A simple instance of this is the bacteria growth models you study in calculus. There, N_t is the number of bacterial in a dish at time t, which is modeled as satisfying the differential equation $\frac{d}{dt}N = gN$, where g is a growth rate coefficient. Even though the actual number of bacteria at any given time is an integer, there is little to gain by distinguishing between $N = 1,000,000$ and $N = 1,000,001$, or the obviously inexact but very accurate $N = 1,000,000.43$. The point is that if N is large and if goes up and down by single one unit at a time, it makes sense to model N_t as a continuously varying even though that is not exactly true.

This is the informal way to make an SDE model of the queue with abandonment system. We will discuss a more formal systematic approach through *scaling* at some point. You just compute $E[\Delta N \mid \mathcal{F}_t]$ and $E[(\Delta N)^2 \mid \mathcal{F}_t]$. Then you look at the calculation and identify $a(N)$ and $b(N)$. In the present case,

$$
\begin{aligned}
E[\,\Delta N \mid \mathcal{F}_t] &= (+1)\cdot \Pr(N \to N+1) \,+\, (-1)\cdot \Pr(N \to N-1) \\
&= \lambda dt \,-\, (\mu + N_t r)\,dt \\
&= (\lambda - \mu - rN_t)\,dt\;.
\end{aligned}
$$

The variance calculation is easier still. The only possible non-zero value of $(\Delta N)^2$ is $(\Delta N)^2 = 1$, and that has probability $(\lambda + \mu + rN_t)\,dt$. Therefore

$$
E[\,\Delta N \mid \mathcal{F}_t] \;=\; (\lambda + \mu + rN_t)\,dt\;.
$$

Therefore, if there is a valid SDE model, it is

$$
dN_t \;=\; (\lambda - \mu - rN_t)\,dt \,+\, (\lambda + \mu + rN_t)^{1/2}\,dW_t\;. \tag{8}
$$

This model is valid, if at all, when N_t is large. This is why we did not bother to think about what happens when $N_t = 0$ and there are no service events. To stress this point yet again, the SDE (8) refers to a Brownian motion W_t, but that is purely for convenience. The derivation had nothing to do with Brownian motion. The dW_t term in (8) just stands for a random variable, the noise, that (a) is non-anticipating, (b) has mean zero, (c) has variance dt, (d) has continuous paths. It turns out that any such process is equivalent to Brownian motion, but that fact is not really relevant for modeling.

Here is another calculation that illustrates the power of the weak side.[1] Suppose $W_{k,t}$, for $k = 1, \ldots, n$ are independent standard Brownian motion paths. We can form the vector W_t whose components are the $W_{k,t}$. This is a standard $n-$dimensional Brownian motion, much in the way $Z = (Z_1, \ldots Z_n)$ is an n component multivariate normal with independent components $Z_k \sim \mathcal{N}(0,1)$. The standard multivariate normal in n dimensions is $Z \sim \mathcal{N}(0, I)$. Its components are independent one dimensional standard normals. The *Bessel process* is the stochastic process

$$
R_t \;=\; |W_t| \;=\; \left(\sum_{k=1}^{n} W_{k,t}^2 \right)^{1/2}\;. \tag{9}
$$

We are going to write the *SDE* satisfied by R_t in the weak sense by calculating $E[\,dR \mid \mathcal{F}_t]$ and $E\Big[\,(dR)^2 \mid \mathcal{F}_t\Big]$.

[1] This is a reference to the *dark side* in Star Wars movies.

The main trick to this calculation is the n variable version of Ito's lemma. Suppose W_t is a standard n component Brownian motion as above and we have $X_t = u(W,t)$, where $u(w,t) = u(w_1,\ldots,w_n,t)$ is a differentiable function of its arguments. The answer here, as before, should depend on the first partial derivatives of u and the second partials with respect to the w variables. It should be enough to start with

$$
\begin{aligned}
du(W_t,t) \;=\; & \sum_{k=1}^{n} \left[\, \partial_{w_k} u(W_t,t)\right] dW_{k,t} \\
& + \left[\, \partial_t u(w,t)\right] dt \\
& + \frac{1}{2}\sum_{k=1}^{n} \left[\, \partial_{w_k}^2 u(W_t,t)\right] \left(dW_{k,t}\right)^2 \\
& + \sum_{j<k} \left[\, \partial_{w_k}\partial_{w_j} u(W_t,t)\right] \left(dW_{j,t} dW_{k,t}\right) \; .
\end{aligned}
$$

On the third line on the right, we should replace $(dW_{k,t})^2$ by dt because $E\Big[\,(dW_{k,t})^2\Big] = dt$, and the variance is too small to matter. We went over this in detail in the one Brownian motion discussion, and it's still true now. In the fourth and last line on the right, we should replace $(dW_{j,t}dW_{k,t})$ by 0, because $E[\,(dW_{j,t}dW_{k,t})] = 0$ (because W_j and W_k are independent for $j \neq k$), and the variance (which is the same size as for $(dW_{k,t})^2$ before), again is too small to matter. The Ito's lemma is

$$
du(W_t,t) \;=\; \sum_{k=1}^{n} \left[\, \partial_{w_k} u(W_t,t)\right] dW_{k,t} \;+\; \left\{ \partial_t u(w,t) \;+\; \frac{1}{2}\sum_{k=1}^{n} \partial_{w_k}^2 u(W_t,t) \right\} dt \tag{10}
$$

It is convenient to write this in vector/operator notation. The gradient of u with respect to the vector w is, $\nabla_w u$. Its components are $\left[\nabla_w u\right]_k = \partial_{w_k} u$. The first sum on the left is the dot product of $\nabla_w u$ with the vector dW_t. That is, $\sum_{k=1}^{n}\left[\,\partial_{w_k} u(W_t,t)\right] dW_{k,t} \;=\; \nabla_w u(W_t,t)\, dW_t$. The last term on the right involves the *Laplace operator*, or the *Laplacian* applied to u. The definition is

$$
\triangle_w u(w,t) \;=\; \sum_{k=1}^{n} \partial_{w_k}^2 u(w,t) \; .
$$

This leads to the more compact statement

$$
du(W_t,t) \;=\; \nabla_w u(W_t,t)\, dW_t \;+\; \left(\partial_t u(W_t,t) \;+\; \frac{1}{2} \triangle_w \, u(W_t,t) \right) dt \; . \tag{11}
$$

This is Ito's lemma for multi-dimensional Brownian motion.

To apply (11) to the Bessel process (9) we have to compute ∇u and $\triangle u$, where $u(w) = |w| = \left(\sum_k w_k^2\right)^{1/2}$. The first derivative is an application of the chain rule:

$$\partial_{w_1} \left(\sum_k w_k^2 \right)^{1/2} = \frac{1}{2} \left(\sum_k w_k^2 \right)^{-1/2} \cdot 2w_1 = w_1 \left(\sum_k w_k^2 \right)^{-1/2} .$$

The second derivative is (with the product rule and the chain rule)

$$\begin{aligned}
\partial_{w_1}^2 u &= \partial_{w_1} \left[w_1 \left(\sum_k w_k^2 \right)^{-1/2} \right] \\
&= \left(\sum_k w_k^2 \right)^{-1/2} + w_1 \cdot 2w_1 \left(\frac{-1}{2} \right) \left(\sum_k w_k^2 \right)^{-3/2} \\
&= \frac{1}{|w|} - \frac{w_1^2}{|w|^3} .
\end{aligned}$$

Summing over k gives

$$\triangle |w| = \sum_{k=1}^n \partial_{w_k}^2 |w| = \frac{1}{|w|} \left[\sum_{k=1}^n \left(1 - \frac{w_k^2}{|w|^2} \right) \right] = \frac{1}{|w|} \left[n - \frac{\sum_{k=1}^n w_k^2}{|w|^2} \right] = \frac{n-1}{|w|} .$$

When we put these calculations into the general formula (11), the result is

$$dR_t = d|W_t| = \frac{1}{|W_t|} W_t \, dW_t + \frac{1}{2} \frac{n-1}{|W_t|} \, dt . \tag{12}$$

The purpose of all this algebra was the calculation of $E[dR]$ and $E[dR^2]$. When we calculate $E[dR_t \mid \mathcal{F}_t]$ we use the fact that W_t is known in $\mathcal{F}_t$ and $E[dW_t \mid \mathcal{F}_t] = 0$. Therefore,

$$E[\, dR_t \mid \mathcal{F}_t] = \frac{1}{2} \frac{n-1}{R_t} \, dt .$$

For the variance calculation, we ignore the drift part just used. Then we calculate

$$E\left[\left| \frac{1}{|W_t|} W_t \, dW_t \right|^2 \right] = \frac{1}{R_t^2} \sum_{k=1}^n W_{k,t}^2 E\Big[(dW_{k,t})^2 \mid \mathcal{F}_t \Big] = \frac{1}{R_t^2} \sum_{k=1}^n W_{k,t}^2 \, dt = dt .$$

You might think there is a shorter way to get a result this simple. There is, but it would take me as long to explain it as this calculation. However you get there, the result is

$$dR_t = \frac{n-1}{2} \frac{1}{R_t} \, dt + dW_t . \tag{13}$$

Warning: the W_t here is a one dimensional Brownian motion. In the weak formulation, we do not try to relate it to the n dimensional Brownian motion used to define the Bessel process.

Look for a moment at the terms in (13). The noise term for R_t is the same as for ordinary Brownian motion. The drift term is positive. It expresses the fact that a multi-dimensional Brownian motion is more likely to move away from the origin than toward the origin. As a check on the calculation, note that this outward "force" vanishes in one dimension as it should. The outward force becomes infinite as $R \to 0$ reflecting the fact that it becomes increasingly hard to make progress toward the origin as you get close. The coefficient $n-1$ has a big impact on the behavior near zero. For $n > 2$ (non-integer values of n are allowed here) the probability is zero for R_t ever to reach zero. That is, in dimensions 3 and higher, the multivariate Brownian motion never touches the origin (if it does not start there). For $n = 2$ the situation is completely reversed. Indeed, $R_t = 0$ for some $t > 0$ almost surely.

You can understand the outward drift in the following geometrical way. Think of the circle (or sphere, if $n > 2$) of radius R_t. Ask whether $W_{t+\Delta t}$ (the multi-variate Brownian motion) is more likely to be inside or outside of this sphere. The sphere curves toward the origin, so the volume near W_t with $r < R_t$ is slightly less than the volume with $r > R_t$. For this reason, $R_{t+\Delta t} = |W_{t+\Delta t}| > R_t$ is slightly more likely than $R_{t+\Delta t} < R_t$. The outward "force" arises from the geometry of spheres rather than a physical force.