LECTURE 7

Diffusions and stochastic differential equations. A diffusion X_t on $\mathbf{R}^d$ is a Markov process with continuous paths which 'locally' looks like Brownian motion W_t dilated by a factor $\sigma(X_t,t)$ plus a 'drift' $b(X_t,t)$. It is written formally as a stochastic differential equation

$$dX_t = \sigma(X_t,t)dW_t + b(X_t,t)dt.$$

The easiest case is when σ and b are just constants in which case the answer is $X_t = \sigma W_t + bt$. Note that in more than one dimension, σ is a matrix and b is a vector.

You can think of the equation as some sort of limit of

$$X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}} = \sigma(X_{\frac{j}{2^n}}, \tfrac{j}{2^n})(W_{\frac{j+1}{2^n}} - W_{\frac{j}{2^n}}) + b(X_{\frac{j}{2^n}}, \tfrac{j}{2^n})\frac{1}{2^n}.$$

Note the right hand side is, given $X_{\frac{j}{2^n}}$, just a normal mean $b(X_{\frac{j}{2^n}}, \frac{j}{2^n})$ and variance $a(X_{\frac{j}{2^n}}, \frac{j}{2^n})$ where $a = \sigma^T\sigma$.

Stochastic integrals. The stochastic differential equation doesn't make ordinary calculus sense because W_t is not differentiable. What it means is

$$X_t - X_0 = \int_0^t \sigma(X_s,s)dW_s + \int_0^t b(X_s,s)ds$$

where the second integral is just a normal integral, but the first integral is a stochastic integral defined by

$$\int_0^t \sigma(X_s,s)dW_s = \lim_{n\to\infty} \sum_{j=0}^{\lfloor 2^n t\rfloor -1} \sigma(X_{\frac{j}{2^n}}, \tfrac{j}{2^n})(W_{\frac{j+1}{2^n}} - W_{\frac{j}{2^n}}).$$

Note that the integrand is always evaluated at the left hand end point or in other words, the Brownian increments always look into the future. Therefore if X_t is adapted to the Brownian filtration $\mathcal{F}_t$ we have that

$$\int_0^t \sigma(X_s,s)dW_s \quad \text{is a martingale.}$$

The stochastic integral can also be defined for σ depending on ω through more than just X_s.

$$\int_0^t \sigma(\omega,s)dW_s = \lim_{n\to\infty} \sum_{j=0}^{\lfloor 2^n t\rfloor -1} \sigma(\omega, \tfrac{j}{2^n})(W_{\frac{j+1}{2^n}} - W_{\frac{j}{2^n}}).$$

If you want this to be a martingale it is clear that you want $\sigma(\omega,t)$ to be measurable with respect to $\mathcal{F}_t$ for each $t \geq 0$. This is called *non-anticipating* (there are other words like *progressively measurable* and *previsible* which mean more or less the same thing).

One of the reasons this is useful is that in fact every continuous martingale arises in this way.

Diffusions are continuous but not necessarily martingales because they have a drift term. So they are martingales + a process of bounded variation, sometimes refered to as *semi-martingales.* They are martingales exactly when the drift $b = 0$. On the other hand not all continuous semi-martingales are diffusions. Diffusions come from the special case where σ only depends on the history of the path through its present value. Or in other words, they are the Markov cases.

Equivalent definitions of diffusions. 1. X_t is a Markov process with transition densities $p(s, x, t, y)$ satisfying, in the t, y variables (with x and s fixed) the *Fokker-Planck* or *forward* equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 (a_{ij} p)}{\partial y_i \partial y_j} - \sum_{i=1}^{d} \frac{\partial (b_i p)}{\partial y_i}$$

for $t > s$ with initial condition $p(s, x, t, y) = \delta_x(y)$ when $t = s$; and in the s, x variables the *Kolmogorov* or *backward* equation

$$\frac{\partial p}{\partial s} = -\frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j} - \sum_{i=1}^{d} b_i \frac{\partial p}{\partial x_i}$$

for $0 \leq s < t$ with final condition $p(s, x, t, y) = \delta_y(x)$ when $s = t$.

1'. For any smooth function f with compact support,

$$f(X_t) - f(X_0) - \int_0^t \Big\{ \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(X_s, s) \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) - \sum_{i=1}^{d} b_i(X_s, s) \frac{\partial f}{\partial x_i}(X_s) \Big\} ds$$

is a martingale.

In either case one says X_t has *generator*

$$L = \frac{1}{2} \sum_{i,j=}^{d} a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i \frac{\partial}{\partial x_i}.$$

2. For any $\lambda \in \mathbf{R}^d$,

$$\exp \Big\{ \lambda \cdot (X_t - X_0 - \int_0^t b(X_s, s) ds) - \frac{1}{2} \int_0^t \lambda^T \cdot a(X_s, s) \lambda ds \Big\}$$

is a martingale. We call a the (infinitesimal) variance and b the (infinitesimal) drift.

Explanations. 1. First lets compute from the Fokker-Planck equation $\lim_{h \to 0} \frac{1}{h} \int f(y - x)p(t, x, t+h, y)dy$. You get $\int f \frac{\partial p}{\partial t} dy$. By the Fokker-Planck equation it is $\int f[\frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 (a_{ij} p)}{\partial y_i \partial y_j} -$

$\sum_{i=1}^d \frac{\partial(b_i p)}{\partial y_i}]dy$. Integrating by parts gives $\int [\frac{1}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2 f}{\partial y_i \partial y_j} + \sum_{i=1}^d b_i \frac{\partial f}{\partial y_i}]p dy$. Choosing $f(x) = x$ and x^2 we get

$$\lim_{h\to 0}\frac{1}{h}\int (y_i - x_i)p(t,x,t+h,y)dy = b_i(x,t)$$
$$\lim_{h\to 0}\frac{1}{h}\int (y_i - x_i)(y_j - x_j)p(t,x,t+h,y)dy = a_{ij}(x,t)$$

or, equivalently,

$$\lim_{h\to 0}\frac{1}{h}E[X^i_{t+h} - X^i_t \mid X_t] = b_i(x,t)$$
$$\lim_{h\to 0}\frac{1}{h}E[(X^i_{t+h} - X^i_t)(X^j_{t+h} - X^j_t) \mid X_t] = a_{ij}(x,t).$$

So it just says that X_t has infinitesimal mean b and variance a.

2. Pick a small h. By the martingale property

$$E\Big[\exp\Big\{\lambda\cdot(X_{t+h} - X_t) - \int_t^{t+h}\lambda\cdot b(X_s,s)ds - \frac{1}{2}\int_t^{t+h}\lambda^T\cdot a(X_s,s)\lambda ds\Big\}\ \Big|\ \mathcal{F}_t\Big] = 1$$

which says that for small h, $X_{t+h} - X_t$ is more or less a multivariate Gaussian with mean $\int_t^{t+h} b(X_s,s)ds$ and covariance $\int_t^{t+h} a(X_s,s)ds$.

Computation of a stochastic integral. For practice let's compute the stochastic integral

$$A_t = \int_0^t W_s dW_s.$$

From the definition we know it should be the limit

$$A_t = \lim_{n\to\infty}\sum_{j=0}^{\lfloor 2^n t\rfloor - 1} W_{\frac{j}{2^n}}(W_{\frac{j+1}{2^n}} - W_{\frac{j}{2^n}}).$$

It's not exactly clear how to do it, but there is a trick. Let's let

$$B_t = \lim_{n\to\infty}\sum_{j=0}^{\lfloor 2^n t\rfloor - 1} W_{\frac{j+1}{2^n}}(W_{\frac{j+1}{2^n}} - W_{\frac{j}{2^n}}).$$

Then $B_t - A_t$ is just the quadratic variation process, which we know is t, and

$$B_t + A_t = \lim_{n\to\infty}\sum_{j=0}^{\lfloor 2^n t\rfloor - 1}(W^2_{\frac{j+1}{2^n}} - W^2_{\frac{j}{2^n}}) = W_t^2.$$

So

$$A_t = \frac{1}{2}([B_t + A_t] - [B_t - A_t]) = \frac{1}{2}(W_t^2 - t).$$

Stochastic calculus. Ordinary calculus is based on the chain rule. The chain rule for diffusions is Ito's formula: If f is a smooth function and $dX_t = \sigma dW_t + bdt$ then (in one dimension)

$$df(X_t) = f'(X_t)dX_t + \frac{\sigma^2}{2}f''(X_t)dt.$$

Of course this is just shorthand for saying that

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\sigma(X_s,s)dW_s + \int_0^t f'(X_s)b(X_s,s)ds + \frac{1}{2}\int_0^t \sigma^2(X_s,s)f''(X_s)ds.$$

Ito's formula is easy to prove. You just write

$$f(X_t) - f(X_0) = \lim_{n\to\infty} \sum_{j=0}^{\lfloor 2^n t\rfloor - 1} [f(X_{\frac{j+1}{2^n}}) - f(X_{\frac{j}{2^n}})].$$

Then you expand in a Taylor series

$$\begin{aligned} f(X_{\frac{j+1}{2^n}}) - f(X_{\frac{j}{2^n}}) =& f'(X_{\frac{j}{2^n}})(X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}}) + \frac{1}{2}f''(X_{\frac{j}{2^n}})(X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}})^2 \\ &+ \frac{1}{2\cdot 3}f'''(X_{\frac{j}{2^n}})(X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}})^3 + \cdots \end{aligned}$$

Now you use the stochastic integral equation for $X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}}$ and let $n \to \infty$. The first term gives no special problem. The third and higher terms can easily be seen to go to zero. The only hard thing is the second term. But if you look, it is almost exactly the same as the computation of the quadratic variation of Brownian motion. This is the origin of the extra 'Ito' term in stochastic calculus. The rules are summarized as

$\times$	dW	dW'	dt
dW	dt	0	0
dW'	0	dt	0
dt	0	0	0

where W and W' are independent Brownian motions.

The multidimensional version is just as easy, and the function f can depend on t:

$$df(X_t,t) = \sigma(X_t,t)\nabla f(X_t,t)dW_t + \left\{\frac{\partial f}{\partial t}(X_t,t) + Lf(X_t,t)\right\}dt.$$

The product formula is a particular case of the above. If $dX_1 = \sigma_1 dW_1 + b_1 dt$ and $dX_2 = \sigma_2 dW_2 + b_2 dt$ then

$$dX_1X_2 = X_1dX_2 + X_2dX_1 + \sigma_1\sigma_2 d(Cov(W_1,W_2)).$$

The last term is a little cumbersome, and there is a nice way to write it. Any continuous submartingale X_t can be written as the difference of a continuous martingale and a continuous, increasing adapted process, called its *compensator*,

$$X_t = M_t - A_t.$$

This is called the *Doob-Meyer decomposition.* Also every square of a martingale is a submartingale by Jensen's inequality. So for every continuous martingale M_t we have a compensator for its square which we call the quadratic variation process $\langle M_t, M_t\rangle$. For example, for Brownian motion $\langle W_t, W_t\rangle = t$. This can be polarized $\langle M_t, N_t\rangle = \frac{1}{4}[\langle M_t + N_t, M_t + N_t\rangle - \langle M_t - N_t, M_t - N_t\rangle]$ and you can see that $M_tN_t - \langle M_t, N_t\rangle$ is always a martingale. So we have the formula

$$dM_tN_t = M_tdN_t + N_tdM_t + d\langle M_t, N_t\rangle.$$

If X_t is the solution of the stochastic differential equation

$$dX_t = \sigma dW_t + bdt$$

then

$$X_t - \int_0^t b(X_s, s)ds$$

is a martingale (since it is a stochastic integral $\int_0^t \sigma dW_s$) and its quadratic variation is

$$\langle X_t, X_t\rangle = \int_0^t a(X_s, s)ds.$$

Example: Exponential Brownian motion. The process

$$X_t = \exp\{\sigma W_t + \mu t\}$$

where σ and μ are constants is often used to model stock prices. Applying Ito's formula to the function $f(x,t) = \exp\{\sigma x + \mu t\}$ we find that

$$dX_t = \sigma X_t dW_t + (\mu + \frac{\sigma^2}{2})dt.$$

It is not a martingale unless $\mu = -\sigma^2/2$.

Cameron-Martin-Girsanov. For each nice $a(x,t)$, $b(x,t)$ and starting point x_0 we have a diffusion which we can think of as a measure $P^{a,b}_{x_0}$ on the space of continuous functions starting at x_0. The question is how these measures are related for different a's and b's. Now if we compute the quadratic variation on $[0,T]$, as we did for Brownian motion, we will find that it is

$$\lim_{n\to\infty} \sum_{j=0}^{\lfloor 2^n t\rfloor - 1} |X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}}|^2 = \int_0^T a(X_s, s)ds$$

with probability one under $P^{a,b}_{x_0}$. That means that if $\tilde{a} \neq a$ then $P^{\tilde{a},b}_{x_0}$ is supported on a *completely different set* than $P^{a,b}_{x_0}$, namely the paths for which the above limit is $\int_0^T \tilde{a}(X_s, s)ds$. We say that the two measures are mutually singular.

On the other hand if we change b to $\tilde{b}$ the new measure we get is *absolutely continuous* with respect to the old one, or *equivalent* meaning that $P_{x_0}^{a,\tilde{b}}(A) > 0$ if and only if $P_{x_0}^{a,b}(A) > 0$ and the Radon-Nikodym derivative of one process on $[0,t]$ with respect to the other has an explicit formula. To keep it simple we will compute only the case where one of them is zero. The general formula can be obtained easily by multiplying the corresponding Radon-Nikodym derivatives. The Cameron-Martin-Girsanov formula is

$$\frac{dP_{x_0}^{a,b}}{dP_{x_0}^{a,0}} = \exp\left\{\int_0^t a^{-1}bdX_s - \frac{1}{2}\int_0^t a^{-1}b^2ds\right\}.$$

Here's why: We'll do it in one dimension so that the notation isn't too bad. What we want to show is if we define a measure

$$Q(A) = \int_A \exp\left\{\int_0^t a^{-1}bdX_s - \frac{1}{2}\int_0^t a^{-1}b^2ds\right\} dP_{x_0}^{a,0} \qquad (1)$$

then Q is a diffusion with parameters a and b, in other words, for each λ,

$$\exp\{\lambda(X_t - X_0 - \int bds) - \frac{\lambda^2}{2}\int_0^t ads\} \qquad (2)$$

is a martingale with respect to Q. Let

$$Y_t = \int_0^t (\lambda + a^{-1}b)dX_s = \int_0^t (\lambda + a^{-1}b)\sigma dW_s.$$

Then we know that for any θ,

$$\exp\{\theta(Y_t - Y_0) - \frac{\theta^2}{2}\int_0^t a(\lambda + a^{-1}b)^2ds\}$$

is a martingale with respect to $P_{x_0}^{a,0}$. Choose $\theta = 1$. Rearranging things a bit and using that $Y_t = \lambda X_t + \int_0^t a^{-1}bdX_s$ we have that

$$\exp\{\lambda(X_t - X_0 - \int bds) - \frac{\lambda^2}{2}\int_0^t ads + \int_0^t a^{-1}bdX_s - \frac{1}{2}\int_0^t a^{-1}b^2ds\}$$

is a martingale with respect to $P_{x_0}^{a,0}$. But that is the same thing as $(1), (2)$.

There is an important point to make here. The Radon-Nikodym derivative is with respect to $P_{x_0}^{a,b}$ so the X_s in the integral is the solution of $dX_s = \sigma dW_t + bdt$ where W_t is Brownian motion. On the other hand with respect to $P_{x_0}^{a,\tilde{b}}$, $dX_s = \sigma d\tilde{W}_t + \tilde{b}dt$ where $\tilde{W}_t$ is Brownian motion. So what gives? The point is that it is not the same thing to be a Brownian motion with respect to $P_{x_0}^{a,\tilde{b}}$ and with respect to $P_{x_0}^{a,b}$. In fact if W_t is a Brownian motion with respect to $P^{a,b}_{x_0}$ then $\tilde{W}_t = W_t + \int_0^t \sigma^{-1}(\tilde{b} - b)ds$ is a Brownian motion with respect to $P^{a,\tilde{b}}_{x_0}$. To see this note that $P^{a,b}_{x_0}$ is the distribution of

$$X_t = \int_0^t \sigma dW_s + \int_0^t b ds$$

and $P^{a,\tilde{b}}_{x_0}$ is the distribution of

$$Y_t = \int_0^t \sigma dW_s + \int_0^t \tilde{b} ds.$$

In other words one obtains Brownian motion as

$$W_s = \int_0^t \sigma^{-1} dX_s + \int_0^t \sigma^{-1} b ds \qquad \text{under} \quad P^{a,b}_{x_0}$$

$$\tilde{W}_s = \int_0^t \sigma^{-1} dX_s + \int_0^t \sigma^{-1} \tilde{b} ds \qquad \text{under} \quad P^{a,\tilde{b}}_{x_0}.$$

Problem. Consider the linear stochastic differential equation

$$dX_t = [A(t)X_t + a(t)]dt + \sigma(t)dW_t, \qquad X_0 = x,$$

where W_t is an r-dimensional Brownian motion independent of the initial vector $x \in \mathbf{R^d}$ and the $d \times d$, $d \times 1$ and $d \times r$ matrices $A(t)$, $a(t)$ and $\sigma(t)$ are non-random.

Show that the solution is given by

$$X_t = \Phi(t)[x + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dW_s]$$

where Φ is the $d \times d$ matrix solution of

$$\dot{\Phi}(t) = A(t)\Phi(t), \qquad \Phi(0) = I.$$

(I is the identity matrix with 1's on the diagonal and 0's elsewhere.)