Stochastic Calculus Notes, Lecture 8

Last modified November 21, 2002

1 Multidimensional diffusions

1.1. the Area integral: Suppose

2 Girsanov's theorem

Girsanov's theorem relates solutions of Ito differential equations with different drifts. It is also an example of an interesting possibility in probability, computing the expected value of one random variable by using a random variable with a different probability measure. In Monte Carlo, this is called "importance sampling", and is used to in making accurate estimates of very small probabilities.

2.1. Probability densities and Lebesgue measure: For Brownian motion, we gave a probability measure but not a probability density. For a simple gaussian random variable $X \sim \mathcal{N}(0,1)$ we instead give a probability density, $u(x) = \sqrt{1}\sqrt{2\pi}e^{-x^2/2}$. This is possible because there already is a measure on the probability space $\Omega = R$, Lebesgue measure. When we write $E[V(X)] = \int V(x)u(x)dx$, the dx refers to integration with respect to Lebesgue measure, which is "uniform measure", $\lambda((a,b)) = b-a$ (here $\lambda(A)$ is the Lebesgue measure of A, applied to $A = (a,b)$). It is also possible to define the "standard normal probability measure", P. This is $P(A) = \int_A u(x)dx$. We then have $E[V(X)] = \int_R V(x)dP(x)$ In abstract probability we describe this situation by saying that the gaussian measure P (possibly written dP) is "absolutely continuous" with respect to Lebesgue measure, λ (possibly written dx). The function $u(x)$ is the density of dP with respect to dx, sometimes written $u(x) = \frac{dP}{dx}$. The formal ratio $\frac{dP}{dx}$ is also called the "Radon Nikodym derivative" of the gaussian measure dP with respect to Lebesgue measure dx.

2.2. The Radon Nikodym derivative: A more abstract version of this situation is that there is a probability space Ω, a $\sigma-$algebra of sets $\mathcal{F}$, and two measures $dP(\omega)$, and $dQ(\omega)$. We will suppose that both are probability measures, though this is not necessary; dQ was Lebesgue measure in the previous paragraph. We say that $L(\omega)$ is the Radon Nikodym derivative of dP with respect to dQ and write $L(\omega) = \frac{dP(\omega)}{dQ(\omega)}$ if $P(A) = \int_{\omega \in A} L(\omega)dQ(\omega)$. We use L here and below (instead of u as above), because the Radon Nikodym derivative is closely related to what statisticians call the "likelihood ratio". The definition of L is the same as saying that for any function, $V(\omega)$,

$$E_P[V(\omega)] = \int_\Omega V(\omega)dP(\omega) = \int_\Omega V(\omega)L(\omega)dQ(\omega) = E_Q[V(\omega)u(\omega)] \; . \qquad (1)$$

Following an earlier custom, we write $E_P[\cdot]$ for expectation with respect to the probability measure P.

2.3. Radon Nikodym derivative as likelihood ratio: If X_0 and X_1 are two random variables with densities $u_0(x)$ and $u_1(x)$, then they have probability measures $dP(x) = u_0(x)dx$ and $dQ(x) = u_1(x)dx$ respectively. Therefore, $L(x) = dP(x)/dQ(x) = u_0(x)dx/u_1(x)dx = u_0(x)/u_1(x)$. The Radon Nikodym derivative is the ratio of the probability densities. Statisticians often call probability densities "likelihoods", particularly when thinking of them as a function of some parameter (the mean, variance, etc.). The ratio of probability densities becomes the "likelihood ratio", L. Though our canceling dx from the numerator and denominator is not rigorous, the formula $L = u_0/u_1$ is easy to check in the integral definition (1), as in the following example.

2.4. Example of one dimensional gaussians: Suppose the measure P corresponds to a standard normal, $u_0(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, and Q to a $\mathcal{N}(\mu, 1)$ random variable, $u_\mu(x) = \frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2}$. The Radon Nikodym derivative dP/dQ is given by $L(x) = u_0(x)/u_\mu(x) = e^{-\mu x+\mu^2/2}$. We can verify this by checking, using the standard gaussian integration formulas for expectation values, that

$$\begin{aligned} E_0[V(X)] &= \frac{1}{\sqrt{2\pi}}\int_R V(x)e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_R V(x)L(x)e^{-(x-\mu)^2/2}dx \\ &= E_\mu[V(X)L(X)] \ . \end{aligned}$$

2.5. Absolutely continuous measures: It might seem that it is easy to calculate the Radon Nikodym derivative, and it generally is, provided it exists. Given probability measures P and Q on the same space, Ω, with the same measureable sets, $\mathcal{F}$, there might be an event, A, that has probability zero in the Q probability but positive P probability. In that case, it is impossible to have

$$P(A) = \int_A L(\omega)dQ(\omega) \ ,$$

because the left side must be zero if $Q(A) = 0$. The Radon Nikodym theorem says that this is the only thing that can go wrong.
Theorem: If $Q(A) = 0$ implies $P(A) = 0$ for every measurable event, A, then there is a Radon Nikodym derivative, $L(\omega)$, that represents P in terms of Q.

2.6. Examples in finite dimensions: These illustrate the possibility that one measure may not be absolutely continuous with respect to another, but they do not give much intuition about the subtlty of absolute continuity applied to diffusions. As one example, consider two one dimensional random variables, the standard exponential and the standard normal. The exponential random

variable has probability density $u_1(x) = 0$ if $x < 0$ and $u(x) = e^{-x}$ for $x > 0$. The standard normal has density $u_0(x)$ as above. The event $A = \{\text{“}x < 0\text{”}\} = (-\infty, 0)$ has Q probability (standard normal probability) $Q(A) = .5$ but P probability $P(A) = 0$, since the exponential random variable is never negative. In this situation we say that the gaussian measure is not absolutely continuous with respect to the exponential measure. On the other hand, the exponential measure is absolutely continuous with respect to gausian measure: an event whose gaussian probability is zero also has exponential probability zero.

As another example, suppose we choose the random variable X by first tossing a “fair” coin. If the toss is H (probability p), we set $X = 0$. If the toss is T (probability $q = 1 - p$), we make X to be a standard normal. The probability density for the resulting random variable is

$$u(x) = p\delta(x) + frac1 - p\sqrt{2\pi}e^{-x^2/2} \ .$$

This density is a “mixture” of the delta function and the standard normal density. It is not absolutely continuous with respect to the “pure” standard normal measure because the event $X = 0$ has probability $p > 0$ for u but probability zero for the standard normal alone. Here, the lack of absolute continuity is caused by a concentration of probability rather than the density being zero in a large region ($x < 0$ above).

2.7. Cantor measure: This shows that it is possible to concentrate probability in a set of Lebesgue measure zero without concentrating it at a point as in the delta measure. The Cantor measure (after Georg Cantor, a late ninteenth century German mathematician) is defined on the interval $0 \le x \le 1$ by throwing out all the “middle thirds” and concentrating the measure on what remains, which is called the Cantor set, C. To determine whether an $x \in [0, 1]$ is in C, we give it's representation base 3: $x = 0.a_1a_2a_3\cdots$, where each a_k is one of the numbers 0, 1, or 2, and

$$x = \sum_{k=1}^{\infty} 3^{-k}a_k \ .$$

The ordinary decimal representation is the same thing with 3 replaced by 10. For example,

$$\frac{1}{4} = 0.020202\cdots \quad , \quad \frac{1}{5} = .012012\cdots \ .$$

The Cantor set is the set of numbers that have $a_k \neq 1$ for all k. The condition $a_1 \neq 1$ rules out all numbers $x \in (\frac{1}{3}, 23)$, the middle third of $(0, 1)$. The middle thirds of the first third and third third are ruled out by the condition $a_2 \neq 0$; numbers of the form $0.01a_3, a_4\cdots$ form the interval $(\frac{1}{9}, \frac{2}{9})$, which is the middle third of the first third, and numbers of the form $0.21a_3, a_4\cdots$ form the interval $(\frac{7}{9}, \frac{8}{9})$, which is the middle third of the third third.

With respect to uniform measure (Lebesgue measure) in the unit interval, the Cantor set has probability zero. The probability that x will be thrown out because $a_1 = 1$ is $\frac{1}{3}$. If x is spared (probability $\frac{2}{3}$), it is thrown out because

$a_2 = 1$ again with probability $\frac{1}{3}$. The probability of being spared k times is $\frac{2}{3}^k \to 0$ as $k \to \infty$. To be in C, x must be spared infinitely many times, an event of Lebesgue measure zero.

To define the Cantor measure, we need to give $P_C(A)$ for any $A \subseteq C$. For each $x \in A$ we define a $y \in (0,1)$ by giving the *base 2* binary expansion $y = 0.b_1, b_2, b_3 \cdots$, where $b_k = 0$ if $a_k = 0$ and $b_k = 1$ if $a_k = 2$. That is

$$y = \sum_{k \in T(x)} 2^{-k} \quad , \quad \text{where } T(x) = \{k \mid a_k = 2\} \quad .$$

The set of all such y coming from an $x \in A$ is called B. The Cantor measure of A will be the ordinary Lebesgue (unifrom) measure of B. If $A \subseteq [0,1]$, we define $P_C(A) = P_C(A \cap C)$. We can think of the Cantor measure as coming from an infinite sequence of cut and squeeze steps. First we cut the unit interval in the middle and squeeze the first half into the first third and the second half to the third third. This gives the first and third thirds probability $\frac{1}{2}$ each and the middle third probability zero. We then cut and squeeze out the middle third of the first and third thirds, giving each of the 4 remaining ninths measure $\frac{1}{4}$, and so on.

The Cantor set and Cantor measure provide illustrate some things that can go wrong in measure theory. . . .

2.8. Alternative descriptions of a random variable: It often happens that we can describe a random $X \in \mathcal{S}$ either as a random variable in its own right by giving a probability measure, Q, on $\mathcal{S}$, or as a function of another random variable $\omega \in \Omega$ with measure P. If we have a function $V(x)$ and we want the expected value, $E[V(X)]$, we may calculate it either as $\int_{\mathcal{S}} V(x) dQ(x)$ or as $\int_\Omega V(X(\omega) dP(\omega)$. Of course, the function $X(\omega)$ and the measure P determine the measure Q. Nevertheless, we sometimes can make use of a direct description of Q without reference to P and Ω. Girsanov's theorem is about the measure in path space defined by the solution of a stochastic differential equation. In this case, Ω is the space of Brownian motion paths and $X(W)$ is the solution of the SDE for Brownian motion path W, which plays the role of ω here. To state Girsanov's theorem, we have to be able to understand the X measure without reference to the underlying W.

2.9. A one dimensional mapping example: Suppose $\Omega = (0,1]$ and P is uniform measure, leaving out the point zero for simplicity. For each $\omega \in (0,1]$ we define $X(\omega) = -\ln(\omega)$ ($\ln(\omega)$ is the log base e). This is a $1-1$ transformation; there is a unique $X \geq 0$ for each $\omega \in (0,1]$, and vice versa. If $V(x) = x^2$, we could evaluate $E[V(X)]$ as the integral $\int_0^1 \ln(\omega)^2 d\omega$.

The other way is to find the PDF for X directly. Since $X \geq 0$, this density is zero for $x < 0$. We call it $u(x)$ and find it from the relation

$$\begin{aligned} u(x)dx &= P(x \leq X \leq x + dx) \\ &= P(x < -\ln(\omega) < x + dx) \end{aligned}$$

$$
\begin{aligned}
&= P(-x - dx < \ln(\omega) < -x) \\
&= P(e^{-x}e^{-dx} < \omega < e^{-x}) \\
&= P(e^{-x} - dxe^{-x} < \omega < e^{-x}) \\
&= dxe^{-x} \ .
\end{aligned}
$$

The last line is because ω is uniformly distributed in $(0, 1]$ so the probability of being in any interval (a, b) is $b-a$, here with $a = e^{-x} - dxe^{-x}$ and $b = e^{-x}$. The conclusion is that $u(x) = e^{-x}$, which is to say that X is a standard exponential random variable. Now we can calulate the same expected value as

$$E[X^2] = \int_{x=0}^{\infty} x^2 e^{-x} dx = \int_{\omega=0}^{1} \ln(\omega)^2 d\omega \ .$$

The P measure is uniform measure on $(0, 1]$. The Q measure is standard exponential measure. The mapping is $X(\omega) = -\ln(\omega)$.

2.10. Distinguishing random variables: Suppose we have probability measures P and Q on the same space, $\mathcal{S}$. A "sample" will be a random variable $X \in \mathcal{S}$ with either the P or Q probability measures. One of the main questions in statistics is finding statistical tests to determine whether X was drawn from the P or Q populations, i.e., which of P or Q describes X. A "hypothesis test" is a decomposition of $\mathcal{S}$ into two sets, here called A_P and A_Q (A_P and A_Q disjoint, $A_P \cup A_Q = \mathcal{S}$). The hypothesis test based on this decomposition reports $X \sim P$ if $X \in A_P$ and $X \sim Q$ if $X \in A_Q$. Generally speaking, in statistics your hypothesis test conclusions are not certain, but hold with a certain (hopefully high) likelihood.

Suppose there is an event $A \subseteq \mathcal{S}$ so that $P(A) > 0$ but $Q(A) = 0$. If we use this set for the hypothesis test, (taking $A_P = A$ and $A_Q = \mathcal{S} - A$), then whenever our hypothesis test reports P, it must be correct. In statisticians' language, there is a hypothesis test with zero type II error. This shows that the possibility of an (in some respectes) infallable hypothesis test is equivalent to absolute continuity or lack of absolute continuity of measures. If there is a procedure that sometimes knows with 100% certainty that X was drawn from Q rather than P, then Q is not absolutely continuous with respect to P. For example, suppose $\mathcal{S} = \mathcal{R}$, Q is the standard normal measure, P is the exponential, and $A = (-\infty, 0)$. If $X \in A$ we know X came from the gaussian measure because the exponential probability of A is zero.

If measure Q is not absolutely continuous with respect to measure P, it is common that the two measures are "completely singular" with respect to each other. This means that there is a partition with $P(A_P) = Q(A_Q) = 1$, and therefore $Q(A_P) = 0 = P(A_Q)$. If measures P and Q are completely singular then there is a hypothesis test that is right 100% of the time. For example, if P is the standard normal measure and $Q = \delta$ is a delta measure corresponding to $X = 0$ with probability 1, then we take A_P to be all real numbers except zero and $A_Q = \{0\}$. The hypothesis test is to say "normal" if $X \neq 0$ and "δ" if $X = 0$. If the only choices are standard normal or delta, you can never be wrong doing this.

2.11. Absolute continuity of diffusion measures: It is possible to distinguish diffusions with different σ in this way. The main fact is that $\langle X \rangle_T = \int_0^T \sigma(X_t, t)^2 dt$. If we know everything about the path, we will be able to compute ($\Delta t = T/n$, $t_k = k\Delta t$):

$$\langle X \rangle_T = \lim_{\Delta t \to 0} \sum_{k=0}^{n-1} \left(X_{t_{k+1}} - X_{t_k} \right)^2 \tag{2}$$

Suppose we are trying to guess whether X satisfies $dX = a_0(X_t, t)dt + \sigma_0(X_t, t)dW_t$ or $dX = a_1(X_t, t)dt + \sigma_1(X_t, t)dW_t$. We compute the quadratic variation for our path X_t (2) and see whether it is equal to $\int_0^T \sigma_0(X_t, t)^2 dt$ or $\int_0^T \sigma_1(X_t, t)^2 dt$. If $\sigma_0^2 \neq \sigma_1^2$, it is impossible for both to be correct (but for the unlikely possiblity that $\sigma_0 = -\sigma_1$). This proves the negative part of Girsanov's theorem:
Theorem: If $\sigma_0^2 \neq \sigma_1$, then the measures corresponding to stochastic processes $dX = a_0 dt + \sigma_0 dW$ and $dX = a_1 dt + \sigma_1 dW$ are completely singular with respect to each other.

2.12. Likelihood ratio for SDE solutions: The positive part of Girsanov's theorem is about the measures for SDE solutions with the same σ but different drift, a. Suppose we have $dX = a_0(X, t) + \sigma(X, t)dW_t$ and $dX = a_1(X, t) + \sigma(X, t)dW_t$, which determine measures in path space $dP_0(X)$ and $dP_1(X)$ respectively. The theorem states that the two measures are absolutely continuous with respect to each other and gives a formula, the Girsanov formula, for the likelihood ratio, or Radon Nikodym derivative, $L(X)$. To be technically more precise, we consider a time T and paths up to time T. Then $L(X)$ is a function of the whole path up to time T.

Our strategy for finding the Girsanov formula is to find the formula for $\overline{L}_{\Delta t}$, the likelihood ratio for the forward Euler approximations to the two processes. The limit $L(X) = \lim_{\Delta t \to 0} \overline{L}_{\Delta t}$ will then be clear. There are probably some technicalities needed for a complete mathematical proof, but I will not dwell on them (an understatement).

2.13. Multiplying conditional probability densities: This is a reminder of rule of conditional probability density density that will be used in the following paragraph. Suppose we first choose a random variable, X, from the probability density $u(x)$, then choose the random variable Y from the conditional density $v_1(y \mid X)$. The resulting pair (X, Y) has joint PDF $U(x, y) = u(x)v(y \mid x)$. If we then choose Z from the density $v_2(z \mid Y)$, the PDF for the triple (X, Y, Z) will be $U(x, y, z) = u(x)v_1(y \mid x)v + 2(z \mid y)$. This is a version of a rule we used for Markov chains: we multiply the conditional probabilities (transition probabilities) to get the joint probability of the path. Here the path is the triple (X, Y, Z), but clearly it could be longer.

2.14. Measure for the forward Euler method: Our standard notation is $\Delta t = T/n$, $t_k = k\Delta t$, $\Delta W_k = W_{k+1} - W_k$, and $\overline{X}_k \approx X_{t_k}$. The approximation

is

$$\overline{X}_{k+1} = \overline{X}_k + a(\overline{X}_k, t_k)\Delta t + \sigma(\overline{X}_k, t_k)\Delta W_k \; . \tag{3}$$

We want an expression for the joint PDF, $U(x_1, \ldots, x_n)$, of the n random variables $\overline{X}_1$, ..., $\overline{X}_n$. The conditional probability density for $\overline{X}_k + 1$ conditioned on $\mathcal{F}_k$ actually depends only on $\overline{X}_k$ and t_k. We call it $u(x_{k+1} \mid x_k; t_k, \Delta t)$. The semicolon separates the conditioning variable, x_k, from the other arguments, t_k and Δt. As in the previous paragraph, this is built up by multiplying the conditional probability densities:

$$U(x_1, \ldots, x_n) = \prod_{k=0}^{n-1} u(x_{k+1} \mid x_k, t_k, \Delta t) \; . \tag{4}$$

For simplicity we suppose that $X_0 = x_0$ is specified and not random. If X_0 were random with probability density $u_0(x_0)$, we would multiply U by this factor.

The big PDF, U, is the PDF for the approximate path, $\overline{X}$, up to time T. Because time is discrete, this approximate path consists of the n values, x_k. We follow the convention of using lower case letters, x_k to be the variables corresponding to the random variables X_k. Suppose, for example, we want to know the probability that the approximate path is less than r for all $t_k \leq T$. This is the event $A = \{x_k \leq r, 1 \leq k \leq n\}$, so it's probability is given by

$$\int_{x_1=-\infty}^{x_1=r} \cdots \int_{x_n=-\infty}^{x_n=r} U(x_1, \ldots, x_n) dx_1 \cdots dx_n = \int_A U(\vec{x}) d\vec{x} \; ,$$

using the notation $\vec{x} = (x_1, \ldots, x_n) \in R^n$ for the n numbers representing the discrete path.

Particularly when Δt is small, the X_k will be strongly correlated with each other, though none is entirely determined by the others. It does not make sense to say the x_k are correlated because they are not random variables. It is true, as we are about to see, that U is very small if x_k is far from the neighboring values x_{k-1} and x_{k+1}, corresponding to the fact that is is unlikely for X_k to be far from the values X_{k-1} and X_{k+1}.

After this buildup, here is the calculation. From the forward Euler formula (3), it is clear that conditioned on $\overline{X}_k$, $\overline{X}_{k+1}$ is a gaussian random variable with mean $\overline{X}_k + a(\overline{X}_k, t_k)\Delta t$ and variance $\sigma(\overline{X}_k, t_k)^2 \Delta t$. Conditioned on $\mathcal{F}_k$, the only thing random in (3) is ΔW_k, which is gaussian with mean zero and variance Δt. Therefore, the formula for $u(x_{k+1} \mid x_k; t_k, \Delta t)$ is just the gaussian formula for x_{k+1} with the appropriate mean and variance:

$$u(x_{k+1} \mid x_k; t_k, \Delta t) = \frac{1}{\sqrt{2\pi\sigma(x_k, t_k)^2 \Delta t}} \exp\left(\frac{-(x_{k+1} - x_k - \Delta t a(x_k, t_k))^2}{2\sigma(x_k, t_k)^2 \Delta t} \right) \; .$$

When we multiply these together we get the desired formula:

$$U(x_1, \ldots, x_n) = \frac{1}{(2\pi\Delta t)^{n/2}} \prod_{k=0}^{n-1} \left(\frac{1}{|\sigma(x_k, t_k)|} \exp\left(\frac{-(x_{k+1} - x_k - \Delta t a(x_k, t_k))^2}{2\sigma(x_k, t_k)^2 \Delta t} \right) \right) \; . \tag{5}$$

2.15. The likelihood ratio: We compute the likelihood ratio $L(\vec{x} = U_0(\vec{x}/U_1(\vec{x}$ the measures given by (5) with drift terms $a_0(x,t)$ and $a_1(x,t)$ respectively. The limit of L when $\Delta t \to 0$ will be clear. We have already seen that we should not expect a reasonable limit if the volatility coefficients are different. This is why we assume that σ is the same for both processes. When taking the ratio, the factors $\sqrt{2\pi\Delta t\sigma^2(x_k,t_k)}$ cancel, leaving only the exponential parts. The result is

$$L(\vec{x}) = \exp\left(\sum_{k=0}^{n-1} \frac{-(x_{k+1} - x_k - a_0(x_k,t_k)\Delta t)^2 + (x_{k+1} - x_k - a_1(x_k,t_k)\Delta t)^2}{2\Delta t\sigma^2(x_k,t_k)}\right) \quad .$$

If we expand the squares as

$$(x_{k+1}-x_k-a_0(x_k,t_k)\Delta t)^2 = (x_{k+1}-x_k)^2-2(x_{k+1}-x_k)a_0(x_k,t_k)\Delta t+a_0(x_k,t_k)^2\Delta t^2$$

and cancel common the common term, there remains

$$\begin{aligned} L(\vec{x}) &= \exp\left(\sum_{k=0}^{n-1} \frac{-(x_{k+1} - x_k)(a_0(x_k,t_k) - a_1(x_k,t_k))}{\sigma^2(x_k,t_k)}\right) \\ &\quad \times \exp\left(\frac{\Delta t}{2}\sum_{k=0}^{n-1} \frac{-(a_0(x_k,t_k) - a_1(x_k,t_k))^2}{2\sigma^2(x_k,t_k)}\right) \end{aligned} \tag{6}$$