Week 10
Change of measure, Girsanov formula

Jonathan Goodman

November 26, 2012

1 Introduction to the material for the week

sec:intro

In Week 9 we made a distinction between *simulation* and *Monte Carlo*. The difference is that in Monte Carlo you are computing a number, A, that is not random. It is likely that there is more than one formula for A. There may be more than one way to express A as the expected value of a random variable. Suppose

$$A = \mathrm{E}[\, F(X)] \ ,$$

where X has probability density $u(x)$. Suppose $v(x)$ is another probability density so that

$$L(x) = \frac{u(x)}{v(x)} \tag{1}$$

eq:L

is well defined. Then

$$A = \int F(x)u(x)dx = \int F(x)\frac{u(x)}{v(x)}v(x)dx \ .$$

This may be written as

$$A = \mathrm{E}_u[\, F(X)] = \mathrm{E}_v[\, F(X)L(X)] \ . \tag{2}$$

eq:is

This means that there are two distinct ways to evaluate A: (i) take samples $X \sim u$ and evaluate F, or (ii) take samples $X \sim v$ and evaluate FL.

Importance sampling means using the *change of measure* formula (2) for Monte Carlo. The expected value $\mathrm{E}_u[\, F(X)]$ means integrate with respect to the *probability measure* $u(x)dx$. Using the measure $v(x)dx$ instead represents a change of measure. The answer A does not change as long as you put the *likelihood ratio* into the second integral, as in the identity (2).

There are many uses of importance sampling in Monte Carlo and applied probability. One use is *variance reduction*. The variance of the $u-$estimator is

$$\mathrm{var}_u(F(X)) = \mathrm{E}_u\big[\, F(X)^2\big] - A^2 \ .$$

The variance of the $v-$estimator is

$$\mathrm{var}_v(F(X)L(X)) = \mathrm{E}_v\Big[\,(F(X)L(X))^2\Big] - A^2\ .$$

It may be possible to find a change of measure and corresponding likelihood ratio so that the variance of the $v-$estimator is smaller. That would mean that the variation of $F(x)L(x)$ is smaller than the variation of $F(x)$, at least in regions that "count". A good probability density v is one that puts more of the probability in regions that are important for the integral, hence the term importance sampling.

Rare event simulation offers especially dramatic variance reductions. This is when $A = \mathrm{P}_u(X \in B)$ (which is the same as $\mathrm{P}_u(B)$) The event B is rare when the probability is small. Applications call for evaluating probabilities ranging from 1% to 10^{-6} or smaller. A good change of measure is one that puts its weight on the most likely parts of B. Consider the one dimensional example where $u = \mathcal{N}(0,1)$ and $B = \{x > b\}$. If b is large, $\mathrm{P}_{0,1}(X > b)$ is very small. But most of the samples with $X > b$ are only a little larger than b. The measure $v = \mathcal{N}(b,1)$ is a simple way to put most of the weight near b. The likelihood ratio is

$$L(x) = \frac{u(x)}{v(x)} = \frac{e^{-x^2/2}}{e^{-(x-b)^2/2}} = e^{b^2/2}e^{-bx}\ .$$

When $x = b$, the likelihood ratio is $L(b) = e^{-b^2/2}$, which is very small when b is large. This is the largest value of $L(x)$ when $x \geq b$.

The probability measures $u(x)dx$ and $v(x)dx$ give two ways to estimate $\mathrm{P}_{0,1}(X > b)$. The $u-$method is to draw L independent samples $X_k \sim \mathcal{N}(0,1)$ and count the number of those with $X_k > b$. Most of the samples are wasted in the sense that they are not counted. The $v-$method is to draw L independent samples $X_k \sim \mathcal{N}(b,1)$. The estimator is

$$\mathrm{P}_{0,1}(X > b) = \int_b^\infty L(x)v(x)dx \approx \frac{1}{L}\sum_{X_k > b} L(X_k)\ .$$

Now about half of the samples are counted. But they are counted with a small weight $L(X_k) < e^{-b^2/2}$. A *hit* is a sample $X_k > b$. A lot of small weight hits give a lower variance estimator than a few large weight hits.

The Girsanov theorem describes change of measure for diffusion processes. Probability distributions, or probability measures, on path space do not have probability densities. *In some cases* the likelihood ratio $L(x)$ can be well defined even when the probability densities $u(x)$ and $v(x)$ are not. If x is a path, the likelihood ratio $L(x)$ is a path function that makes (2) true for "well behaved" functions F. Roughly speaking, a change of measure can change the drift of a diffusion process but not the noise. The *Girsanov formula* is the formula for the L that does the change.

Girsanov's theorem has two parts. One part says when two diffusion processes may be related by a change of measure. If they can be, the two probability

measures are *absolutely continuous* with respect to each other, or *equivalent*. If two probability measures are not equivalent in this sense, then at least one of them has a component that is *singular* with respect to the other. The other part of Girsanov's theorem is a formula for $L(x)$ in cases in which it exists. This makes the theorem useful in practice. We may compute hitting probabilities or expected payouts using any diffusion that is equivalent to the one we are interested.

2 Probability measures

A *probability measure* is function that gives the probability of any event in an appropriate class of events. If B is such an event, then $\mathrm{P}(B)$ is this probability. By "class of appropriate events", we mean a $\sigma-$algebra. A probability function must be *countably additive*, which means that if B_n is a sequence of events with $B_n \subseteq B_{n+1}$ (an *expanding* family of events), then

$$\lim_{n\to\infty} \mathrm{P}(B_n) = \mathrm{P}\left(\bigcup_{n=1}^{\infty} B_n \right) . \tag{3}$$ eq:ca

This formula says that the probability of a set is in some sense a continuous function of the set. The infinite union on the right really is the limit of the sets B_n. Another way to write this is to suppose C_k is any sequence "appropriate events", and define

$$B_n = \bigcup_{k=1}^{n} C_k .$$

Then the B_n are an expanding family. The countable additivity formula is

$$\lim_{n\to\infty} \mathrm{P}\left(\bigcup_{k=1}^{n} C_k \right) = \mathrm{P}\left(\bigcup_{k=1}^{\infty} C_k \right) .$$

Every proof of every theorem in probability theory makes use of countable additivity of probability measures. We do not mention this property very often in this course, which is a signal that we are not giving full proofs.

2.1 Integration with respect to a probability measure

A probability density defines a probability measure. If the probability space is $\Omega = \mathbb{R}^n$ and $u(x)$ is a probability density for an n component random variable $(x_1, \ldots, x_n)$, then

$$\mathrm{P}_u(B) = \int_B u(x)dx .$$

is the corresponding probability measure. If B is a small neighborhood of a specific outcome x, then we write its probability as $\mathrm{P}_u(B) = dP = u(x)dx$.

More generally, if $F(x)$ is a function of the random variable x, then

$$\mathrm{E}_u[\,F(X)] = \int F(x)dP(x)\ . \tag{4}$$ eq:pi

This is the same as $\int F(x)y(x)dx$ when there is a probability density.

But the expression (4) makes sense even when P is a more general probability measure. A simple definition involves a $\Delta F = 2^{-m}$ rather than a Δx. Define the events $B_k^{(\Delta F)}$ as

$$B_k^{(\Delta F)} = \{x \mid k\Delta F \le F(x) < (k+1)\Delta F\}\ . \tag{5}$$ eq:Bk

To picture these sets, suppose x is a one dimensional random variable and consider the graph of a function $F(x)$. Divide the vertical axis into equal intervals of size ΔF and a horizontal line for each breakpoint $k\Delta F$. The set $B_k^{(\Delta F)}$ is the part of the $x-$axis where the graph of F lies in the horizontal stripe between $k\Delta F$ and $(k+1)\Delta F$. This set could consist of several intervals (for example, two intervals if F is quadratic) or something more complicated if F is a complicated function. If Ω is an abstract probability space, then the sets $B_k^{(\Delta F)}$ are abstract events in that space. By definition, the function F is *meeasurable* with respect to the $\sigma-$algebra $\mathcal{F}$ if each of the sets $B_k^{(\Delta F)} \in \mathcal{F}$ for each k and ΔF.

The probability integral (4) is defined as a limit of approximations, just as the Riemann integral and Ito integral are. The approximation in this case is motivated that if $x \in B_k^{(\Delta F)}$, then $|F(x) - k\Delta F| \le \Delta F$. Therefore, if the dP integral were to make sense, we would have

$$\begin{aligned}
&\left| \int_{B_k^{(\Delta F)}} F(x)dP(x) \;-\; k\Delta F \int_{B_k^{(\Delta F)}} dP(x) \right| \\
&\qquad = \left| \int_{B_k^{(\Delta F)}} F(x)dP(x) - k\Delta F\,\mathrm{P}\left(B_k^{(\Delta F)}\right) \right| \\
&\qquad \le \Delta F\,\mathrm{P}\left(B_k^{(\Delta F)}\right)\ .
\end{aligned}$$

The approximation to the integral is defined by using the above approximation on each horizontal slice.

$$\int F(x)dP(x) = \sum_k \int_{B_k^{(\Delta F)}} F(x)dP(x) \approx \sum_k k\Delta F \mathrm{P}\left(B_k^{(\Delta F)}\right)$$

This was motivation. The formal definition of the approximations is

$$I_m = \sum_k k\Delta F \mathrm{P}\left(B_k^{(\Delta F)}\right) \quad , \qquad \text{with } \Delta F = 2^{-m}\ . \tag{6}$$ eq:Li

The probability integral is defined as

$$\int_\Omega F(x)dP(x) = \lim_{m\to\infty} I_m\ . \tag{7}$$ eq:pil

The numbers on the right are a Cauchy sequence because if $n > m$ then

$$\begin{aligned} |I_m - I_n| &\leq \Delta F \sum_k \mathrm{P}\left(B_k^{(\Delta F)}\right) \\ &= 2^{-m} \sum_k \mathrm{P}\left(B_k^{(\Delta F)}\right) \\ &= 2^{-m} \; . \end{aligned}$$

The expected value is the same thing as the probability integral:

$$\mathrm{E}_P[\, F(X)] = \int F(x) dP(x) \; .$$

A different view of our definition of the probability integral will be useful in the next subsection. The *indicator function* of an event B is $\mathbf{1}_B(x) = 1$ if $x \in B$ and $\mathbf{1}_B(x) = 0$ if $x \notin B$. A *simple function* is a finite linear combination of indicator functions. We say $F(x)$ is a simple function if there are events $B_1, \ldots, B_n$ and weights $F_1, \ldots, F_n$ so that

$$F(x) = \sum_{k=1}^{n} F_k \mathbf{1}_{B_k}(x) \; .$$

We could define the probability integral of a simple function as

$$\int F(x) dP(x) = \sum_{k=1}^{n} F_k P(B_k) \; . \tag{8}$$ eq:sfi

This has to be the definition of the integral of a simple function if the integral is linear and if

$$\int_\Omega \mathbf{1}_B(x) dP(x) = \int_B dP(x) = P(B) \; .$$

Once you know what the integral should be for simple functions, you know what it should be for any function that can be approximated by simple functions. If F is a bounded function, then

$$F^{(\Delta F)}(x) = \sum_{k\Delta F < F_{\max}} k\Delta F \mathbf{1}_{B_k^{(\Delta F)}}(x)$$

satisfies $\left|F^{(\Delta F)}(x) - F(x)\right| \leq \Delta F$ for all x. Therefore, if the concept of integration makes sense at all, the following should be true:

$$\begin{aligned} \int_\Omega F(x) dP(x) &= \lim_{\Delta F \to 0} \int_\Omega F^{(\Delta F)}(x) dP(x) \\ &= \sum_{k\Delta F < F_{\max}} k\Delta F \int_\Omega \mathbf{1}_{B_k^{(\Delta F)}} dP(x) \\ \int_\Omega F(x) dP(x) &= \lim_{\Delta F \to 0} \sum_{k\Delta F < F_{\max}} k\Delta F \, P(B_k^{(\Delta F)}) \end{aligned} \tag{9}$$ eq:sfd

The point of this is that if the integral of indicator functions is defined, then all other integrals are automatically defined.

A fully rigorous treatment would stop here to discuss a large number of technical details here. The sum that defines the approximation (6) converges if F is bounded. If F is not bounded, we can "approximate" F by a bounded function and try to take the limit. The most important integration theorem is the *dominated convergence theorem*, which gives a condition under which *pointwise* convergence

$$F_n(x) \to F(x) \text{ as } n \to \infty \text{ almost surely}$$

implies convergence of the probability integrals

$$\int F_n(x)dP(x) \to \int F(x)dP(x) \; . \tag{10}$$ eq:lth

The condition concerns the *maximal function*

$$M_F(x) = \sup_n |F_n(x)| \; .$$

If

$$\int M_F(x)dP(x) < \infty \; , \tag{11}$$ eq:mb

then (10) is true. We do not give the proof in this course. A simple way to show that (11) is satisfied is to come up with a function $G(x)$ so that $|F_n(x)| \leq G(x)$ for all x and all n, and $\int G(x)dP(x) < \infty$. A function G like this is a *dominating* function.

2.2 Absolutely continuity of measures

If P is a probability measure then a function $L(x)$ can define a new measure through the informal relation

$$dQ(x) = L(x)dP(x) \; . \tag{12}$$ eq:dQ

This means that for any measurable set B,

$$Q(B) = \int_B L(x)dP(x) \; . \tag{13}$$ eq:Q

In order for Q to be a probability measure, L must have two properties. First, $L(x) \geq 0$ almost surely (with respect to P). Second,

$$\int_\Omega L(x)dP(x) = 1 \; . \tag{14}$$ eq:L1

If the probability measure P is defined and if L has these two properties, then (13) defines another probability measure Q.

The informal relation (12) leads to a relationship between expectation values in the P and Q measures:

$$\mathrm{E}_Q[\,F(X)] = \mathrm{E}_P[\,L(X)F(X)]\ . \tag{15}$$

eq:EQ

This becomes clearer when you use (12) to derive the equivalent probability integral equation

$$\int_\Omega F(x)dQ(x) = \int_\Omega F(x)L(x)dP(x)\ . \tag{16}$$

eq:piL

You can prove this formula as suggested at the end of the previous subsection. You check that it is true for simple functions. Then it is true for any other function, because any function can be well approximated by simple functions. Moreover, it is true for simple functions if it is true for indicator functions. But if F is an indicator function $F(x) = \mathbf{1}_B(x)$, then (16) is exactly the same as (13).

For probability measures defined by densities, the definition (12) is the same as the original definition (1). If $dP(x) = u(x)dx$ and $dQ = v(x)dx$, then $dQ(x) = \frac{v(x)}{u(x)}dP(x)$.

You can ask the reverse question: given probability measures dP and dQ, is there a function $L(x)$ so that $dQ(x) = L(x)dP(x)$? There is an obvious *necessary condition*, which is that any event that is impossible under P is also impossible under Q. If B is an event with $P(B) = 0$, then the definition (13) gives $Q(B) = 0$. You can see this by writing

$$\int_B L(x)dP(x) = \int_\Omega \mathbf{1}_B(x)L(x)dP(x)\ .$$

If $F(x) = \mathbf{1}(x)L(x)$ and $B_k^{(\Delta F)}$ for this function, then $B_k^{(\Delta F)} \subseteq B$, which implies that $P(B_k^{(\Delta F)}) \leq P(B) = 0$. Therefore, all the approximations to $\int \mathbf{1}_B(x)L(x)dP(x)$ are zero. The *Radon Nikodym theorem* states that this necessary condition is *sufficient*. If P and Q are any two probability measures with the same $\sigma-$algebra $\mathcal{F}$, and if $P(B) = 0$ implies that $Q(B)$, then there is a function $L(x)$ that gives Q from P via (13). This function is called the *Radon Nikodym derivative* of Q with respect to P, and is written

$$L(x) = \frac{dQ(x)}{dP(x)}\ .$$

If the condition $P(B) = 0 \Longrightarrow Q(B) = 0$ is satisfied, we say that Q is *absolutely continuous* with respect to P. This term (absolutely continuous) is equivalent in a special case to something that really could be called absolute continuity. But now the term is applied in this more general context. If P is absolutely continuous with respect to Q and Q is absolutely continuous with respect to P, then the two measures are *equivalent* to each other.

If Q is not absolutely continuous with respect to P, then there is an event B that has positive probability in the Q sense but probability zero in the P sense.

When this happens, it is usual that all of Q has zero probability in the P sense. We say that Q is *completely singular* with respect to P if there is an event B with $Q(B) = 1$ and $P(B) = 0$. If Q is completely singular with respect to P, then P is completely singular with respect to Q, because the event $C = B^c$ has $P(C) = 1$ but $Q(C) = 0$. We write $P \perp Q$ when P and Q are completely singular with respect to each other.

Here is a statistical interpretation of absolute continuity and singularity. Suppose X is a sample from P or Q, and you want to guess whether $X \sim P$ or $X \sim Q$. Of course you could guess. But if your answer is a function of X alone (and not another coin toss), then there is some set $B \subset \Omega$ so that if $X \in B$ you say Q, and otherwise you say P. A *type I error* would be saying Q when the answer is P, and a *type II error* is saying P when the answer is Q.[1] The *confidence* of your procedure is $1 - P(B)$, which is the probability of accepting the null hypothesis P if P is true. The *power* of your test is $Q(B)$, which is the probability of rejecting the null hypothesis when the null hypothesis is false. If $P \perp Q$, then there is a test with 100% confidence and 100% power. You say $X \sim Q$ if $X \in B$ and you say $X \sim P$ otherwise. Conversely, if there is test with 100% confidence and 100% power, then $P \perp Q$.

A statistical test, or, equivalently, a set B, is *efficient* if there is no way to increase the confidence in the test without decreasing its power. Equivalently, B is efficient if you cannot increase its power without decreasing its confidence. The *Neyman Pearson lemma* says that if B is efficient, then there is an L_0 so that $B = \{x \mid L(x) > L_0\}$. Since $L = \frac{dQ}{dP}$, the Q probability is larger than the P probability when L is large.

2.3 Examples

Suppose $X \in \mathbb{R}^n$ is an n component random variable. Suppose P makes X a multivariate normal with mean zero and covariance matrix C. If $H = C^{-1}$, then the probability density for P is

$$u(x) = ce^{-x^t Hx/2} \ .$$

(The normalization constant c is not the covariance matrix C.) We want $L(x) = ce^{y^t x}$ to be a likelihood ratio. What should c be and what is the mean and covariance of the resulting distribution? We find the answer by computing $e^{-(x-\mu)^t H(x-\mu)/2} = e^{-x^t Hx/2} e^{\mu^t Hx} e^{-\mu^t H\mu/2}$. If $\mu^t H = y^t$, then $H\mu = y$ because H is symmetric, so $\mu = Cy$ and $e^{-\mu^t H\mu/2} = e^{-y^t Cy/2}$. Therefore

$$L(x) = e^{y^t x} e^{-y^t Cy/2} \tag{17}$$ eq:LG

is a likelihood ratio, and the Q distribution has the same covariance matrix but mean $\mu = Cy$.

[1] In statistics, P would be the *null hypothesis* and Q the *alternate*. Rejecting the null hypothesis when it is true is a type I error. Accepting the null hypothesis when it is false is a type II error. Conservative statisticians regard type I errors as worse than type II.

Suppose X is a one dimensional random variable, the P distribution is uniform $[0,1]$ and the Q distribution is $\mathcal{N}(0,1)$. Then P is absolutely continuous with respect to Q but Q is not absolutely continuous with respect to P.

Any two Gaussian distributions in the same dimension are equivalent.

Suppose $X \in \mathbb{R}^n$, and $P = \mathcal{N}(0, I)$. Suppose Q is the probability distribution formed by taking $X = \frac{Y}{|Y|}$, were $Y \sim \mathcal{N}(0, I)$. Let $|Y| = (Y_1^2 + \cdots + Y_n^2)^{1/2}$ is the *Euclidean length*. Then X is a unit vector that is uniformly distributed on the unit sphere in n dimensions. That sphere is called S_{n-1}, because it is an $n-1$ dimensional surface. For example, S_2 is the two dimensional unit sphere in three dimensions. The set $B = S_{n-1}$ has $Q(B) = 1$ and $P(B) = 0$.

3 Changing the drift, Girsanov's theorem

The simplest version of *Girsanov's formula* is a formula for the $L(x)$ that changes standard Brownian motion to one with drift a_t. This L relates P, which is the distribution for standard Brownian motion on $[0,T]$ to Brownian motion with drift. That is

$$P: \quad dX_t = dW_t \tag{18}$$ eq:SDEP

$$Q: \quad dX_t = a_t dt + dW_t \; . \tag{19}$$ eq:SDEQ

The formula is $dQ(x) = L(x) dP(x)$, where

$$L(x) = e^{\int_0^T a_t dX_t} e^{-\int_0^T a_t^2 dt/2} \; . \tag{20}$$ eq:G1

The integrals that enter into L are well defined. In fact, we showed that they are defined almost surely. If P and Q are absolutely continuous with respect to each other (are equivalent), then almost surely with respect to P is the same as almost surely with respect to Q. On the other hand, likely with respect to P does not mean likely with respect to Q, as our importance sampling example shows.

There are several ways to derive Girsanov's formula (20). Here is one way that is less slick but more straightforward. Choose a $\Delta t = T 2^{-m}$ and let the *observations* of X at the times $t_k = k\Delta t$ be assembled into a vector $\vec{X} = (X_1, X_2, \ldots, X_{2^m})$. We are writing X_k for X_{t_k} as we have sometimes done before. We write an exact formula for the joint PDF of $\vec{X}$ under P, and an approximate formula for the joint density under Q. The ratio of these has a well defined limit, which turns out to be (20), as $\Delta t \to 0$.

Let $u(\vec{x})$ be the density of $\vec{X}$ under P. We find a formula for u by thinking of a single time step that goes from x_k to x_{k+1}. The conditional density of X_{k+1} given X_k is normal mean zero variance Δt. This makes (in a hopefully clear notation)

$$u_{k+1}(x_1, \ldots, x_{k+1}) = u_k(x_1, \ldots, x_k) \frac{e^{-(x_{k+1} - x_k)^2/(2\Delta t)}}{\sqrt{2\pi \Delta t}} \; .$$

Therefore

$$u(\vec{x}) = u(x_1, \ldots, x_n) = c \, \exp\left(-\sum_{k=0}^{n-1} \frac{(x_{k+1} - x_k)^2}{2\Delta t} \right) . \tag{21}$$ eq:uP

The distribution of $\vec{X}$ under Q can be written (approximately) in a similar way. We call it $v(\vec{x})$. In the Q process, conditional on X_k, X_{k+1} is approximately normal with variance Δt and mean $X_k + a_{t_k}\Delta t$. Therefore,

$$v_{k+1}(x_1, \ldots, x_{k+1}) \approx v_k(x_1, \ldots, x_k) \frac{e^{-(x_{k+1} - x_k - a_{t_k}\Delta t)^2/(2\Delta t)}}{\sqrt{2\pi\Delta t}} .$$

This leads to

$$\begin{aligned} v(\vec{x}) = v(x_1, \ldots, x_n) &= c \, \exp\left(-\sum_{k=0}^{n-1} \frac{(x_{k+1} - x_k - a_{t_k}\Delta t)^2}{2\Delta t} \right) \\ &= c \exp\left(-\sum_{k=0}^{n-1} \frac{(x_{k+1} - x_k)^2}{2\Delta t} \right) \exp\left(\sum_{k=0}^{n-1} (x_{k+1} - x_k) \, a_{t_k} \right) \exp\left(\frac{-\Delta t}{2} \sum_{k=0}^{n-1} a_{t_k}^2 \right) \end{aligned} \tag{22}$$ eq:vQ

Now take the quotient $L = v/u$. The first exponential on the right of (22) eq:vQ cancels. The second in an approximation of an Ito integral

$$\sum_{k=0}^{n-1} (x_{k+1} - x_k) \, a_{t_k} \to \int_0^T a_t dX_t \quad \text{as} \quad \Delta t \to 0 .$$

The third is an approximation to an ordinary integral:

$$\Delta t \sum_{k=0}^{n-1} a_{t_k}^2 \to \int_0^t a_t^2 dt .$$

Therefore

$$\lim_{\Delta t \to 0} \frac{v(\vec{x})}{u(\vec{x})} = \exp\left(\int_0^T a_t dX_t \right) \exp\left(\frac{-1}{2} \int_0^T a_t^2 dt \right) .$$

This is the formula (20). eq:G1

Note the similarity of the Gaussian path space change of measure formula (20) eq:G1 to the simple Gaussian change of measure formula (17). eq:LG In both cases the first exponential has an exponent that is linear x. The second exponential has a quadratic exponent that normalizes L. In (17), eq:LG the first exponent makes Q larger in the direction of y, which is the direction in which $y^t x$ grows the fastest. In (20), eq:G1 the pulling term $\int a_t dX_t$ is large when X_t moves in the direction of a_t. For example, if $a_t > 0$, then (19) eq:SDEQ has a drift to the right, which is direction in which L, given by (20) eq:G1 is large.

The more general Girsanov theorem concerns two stochastic processes

$$P: \quad dX_t = b_1(X_t)dW_t \tag{23}$$ eq:bdW

$$Q: \quad dX_t = a(X_t)dt + b_2(X_t)dW_t \ . \tag{24}$$ eq:abdW

One part of Girsanov's theorem is that P and Q are singular with respect to each other unless $b_1 = b_2$. The proof of this is the quadratic variation formula from an earlier week. If $dX_t = a(X_t)dt + b(X_t)dW_t$, then

$$[X]_t = \int_0^t b(X_s)^2 ds \ . \tag{25}$$ eq:qv

The quadratic variation is

$$[X]_t = \lim_{\Delta t \to 0} \sum_{t_k < t} \left(X_{k+1} - X_k\right)^2 \ . \tag{26}$$ eq:qvd

We showed that the limit exists for a given path X_t almost surely. If you have a path, you can evaluate the quadratic variation by taking the limit (26). If this is gives b_1, then X_t came from the P process (23). If it gives b_2, then X_t came from the Q process (24). Since these are the only choices, you can tell with 100% confidence whether a given path is from P or Q.

This fact has important implications for finance, particularly medium frequency trading. If you have, say, daily return data or ten minute return data, you have a pretty good idea what the volatility is. The higher frequency the measurements, the better your estimate of instantaneous volatility. But taking more frequent measurements does not estimate the drift – expected return – more accurately. With high frequency data, the variance of asset prices, or the covariances of pairs of prices, is more accurately known than the expected returns.

The final part of Girsanov's theorem is a formula for the likelihood ratio between the measures

$$P: \quad dX_t = b(X_t)dW_t \tag{27}$$ eq:bdW1

$$Q: \quad dX_t = a(X_t)dt + b(X_t)dW_t \ . \tag{28}$$ eq:abdW1

This formula can be derived from approximate formulas for the probability density in a way similar to how we got (20). But people prefer the following argument. The change of measure

$$L(w) = e^{\int_0^T \mu_t dW_t} e^{-\int_0^T \mu_t^2 dt/2}$$

Turns the process dW_t into the process $dW'_t = \mu_t dt + dW_t$. Therefore, if we use this weight function, on the process (27) we get

$$dX = b(X_t)\left(\mu_t dt + dW_t\right) \ .$$

This becomes the process (28) if

$$\mu_t = \frac{a_t}{b_t} \ .$$

Therefore, the likelihood ratio between (27) and (28) is

$$L = e^{\int_0^T \frac{a_t}{b_t} dW_t} e^{-\int_0^T \frac{a_t^2}{b_t^2} dt/2}.$$