Week 10
Change of measure, Girsanov

Jonathan Goodman

November 25, 2013

1 Reweighting

Suppose X is a random variable with probability density $u(x)$. Then the expected value of $f(X)$ is

$$\mathrm{E}_u[\, f(X)] = \int f(x)u(x)\, dx \; . \tag{1}$$

Suppose $v(x)$ is a different probability density. The *likelihood ratio* is

$$L(x) = \frac{u(x)}{v(x)} \; . \tag{2}$$

The expectation (1) can be expressed as

$$\begin{aligned} \mathrm{E}_u[\, f(X)] &= \int f(x)u(x)\, dx \\ &= \int f(x)\frac{u(x)}{v(x)}\, v(x)\, dx \\ \mathrm{E}_u[\, f(X)] &= \mathrm{E}_v[\, f(X)L(X)] \; . \end{aligned} \tag{3}$$

This formula represents *reweighting*. The likelihood ratio, $L(x)$ turn u into v. The expectation value is unchanged because L is included in the expectation with respect to v.

This section is about *Girsanov theory*, which is reweighting one diffusion process to get a different diffusion process. The specific topics are:

1. When are two diffusions related by a reweighting? Answer: if they have the same noise. You can adjust the drift by reweighting, but not the noise.

2. If two diffusions are related by a reweighting, what likelihood ratio $L(x_{[1,T]})$? The path $x_{[0,T]}$ is the random variable and L is a function of the random variable. Answer: $L(x_{[0,T]})$ is given in terms of a stochastic integral. We give a direct derivation, which is straightforward but might be considered complicated. Then we show how to verify that the stochastic integral formula is correct if you are presented with it. This is considered (by me) a less useful approach, albeit more common.

3. What are the theoretical and practical uses of reweighting?

1.1 Gaussian examples

You can understand most important things about reweighting using Gaussian examples, including the three items listed above. We go through this doing first a one dimensional Gaussian random variable, then a discrete time Gaussian process, then the Ornstein Uhlenbeck process.

1.1.1 Scalar Gaussian

Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. The two probability densities are

$$X_1 \sim u(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x-\mu_1)^2/2\sigma_1^2}$$
$$X_2 \sim v(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(x-\mu_2)^2/2\sigma_2^2} \; .$$

The likelihood ratio (2) is

$$\frac{u(x)}{v(x)} = \frac{\sigma_2}{\sigma_1} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1^2}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2^2}\right)^2\right]} \; . \tag{4}$$

This answers question 1 above: you can reweight any one dimensional normal to get any other one dimensional normal.

The reweighting expectation formula (3) for one dimensional Gaussians therefore takes the form

$$\begin{aligned} \mathrm{E}_{\mu_1,\sigma_1}[\, f(X)] &= \mathrm{E}_{\mu_2,\sigma_2}[\, f(X)L(X)] \\ &= \frac{\sigma_2}{\sigma_1} \mathrm{E}_{\mu_2,\sigma_2}\left[f(X) e^{-\frac{1}{2}\left[\left(\frac{X-\mu_1}{\sigma_1^2}\right)^2 - \left(\frac{X-\mu_2}{\sigma_2^2}\right)^2\right]} \right] \; . \end{aligned} \tag{5}$$

The case $\sigma_1 = \sigma_2$ is most common. The likelihood ratio formula (4) is much simpler in that case. The reweighting formula (5) simplifies to

$$\mathrm{E}_{\mu_1,\sigma}[\, f(X)] = e^{-\frac{(\mu_1-\mu_2)^2}{2\sigma^2}} \, \mathrm{E}_{\mu_2,\sigma}\left[f(X) e^{\frac{\mu_1-\mu_2}{\sigma^2} X} \right] \; . \tag{6}$$

The simple case $\sigma = 1$, $\mu_2 = 0$, $\mu_1 = \mu$ helps us interpret this formula:

$$\mathrm{E}_{\mu,1}[\, f(X)] = e^{-\mu^2/2} \, \mathrm{E}_{0,1}\left[\, f(X)\, e^{\mu X}\right] \; .$$

The factor $e^{\mu X}$ in the expectation "pulls" the expectation from 0 in the μ direction. If $\mu > 0$, it weights positive X values more and negative X values less. A physical interpretation (for those who have studied statistical physics) is that there is a force, μ, with potential energy $\mathrm{PE} = -\mu X$. Ignoring the temperature,

$e^{\mu X}$ is how we take this into account. The prefactor $e^{-\mu^2/2}$ renormalizes the expression to be a probability distribution. Otherwise, we would get an answer larger than 1 from $f = 1$.

Importance sampling is a Monte Carlo technique that uses reweighting to handle *rare event* problems. A rare event is one that happens with small probability. Direct Monte Carlo simulation is not accurate for rare events because they are rare. Suppose, for instance, that $\mathrm{P}(A) = 10^{-6}$. If you draw $N = 10^5$ Monte Carlo samples, there is a 90% chance none of them are in A. As a result, you learn that $P(A)$ is probably not much larger than 10^{-5}, but you don't know whether A happens once in a hundred thousand, or once in a million, or even less.

In Monte Carlo terminology, you draw *samples*, X_j, from a distribution $u(x)$. A *hit* is a sample $X_j \in A$. The fraction of hits in N samples is a Monte Carlo estimate of $P(A)$. If you draw 10^5 samples and get zero hits, those samples are all wasted. More generally, suppose you want to evaluate $M = \mathrm{E}_u[\, f(X)]$. The vanilla Monte Carlo estimator with N samples is

$$\widehat{M}_u = \frac{1}{N} \sum_{j=1}^{N} f(X_j) \; . \tag{7}$$

The vanilla estimator is inaccurate if $\mathrm{P}_u(f(X) \neq 0)$ is small. We estimate $\mathrm{P}_u(X \in A)$ using the indicator function $f(x) = \mathbf{1}_A(x)$.

Importance sampling means choosing a distribution $v(x)$ so that $\mathrm{P}_v(X \in A)$ is much larger than $\mathrm{P}_u(X \in A)$. The importance sampler estimate of $M = \mathrm{P}_u(X \in A)$ is

$$\widehat{M}_v = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_A(X_j) L(X_j) \; , \tag{8}$$

If v is well chosen, then there are many hits in the v simulation. The estimator is small, not because there are few hits, but because the hits are counted with small weights $L(X_j)$. The variance of the fancy estimator (8) can be much smaller than the variance of the vanilla estimator (7). In the interest of full disclosure, a bad important sampling strategy can make the variance worse. There is no theorem stating that $\mathrm{var}\Big(\widehat{M}_v\Big) \leq \mathrm{var}\Big(\widehat{M}_u\Big)$. Monte Carlo practitioners can tell you stories of the opposite.

For example, suppose $X \sim \mathcal{N}(0,1)$ and we want to know $\mathrm{P}(X > K)$, for some large K. Assignment 10 shows that importance sampling with $v = \mathcal{N}(K,1)$ is much more efficient than vanilla Monte Carlo.

1.1.2 Discrete time Gaussian process

Suppose $X_0 = x_0$ is given and not random, and then

$$X_{k+1} = aX_k + bZ_k \; , \tag{9}$$

where the $Z_k \sim \mathcal{N}(0,1)$ are i.i.d. The path up to time T is $X_{[1:T]} = (X_1, X_2, \ldots, X_T)$. Let $u(x_{[1:T]})$ be the joint PDF of $(X_1, \ldots, X_T)$. Suppose $v(x_{[1:T]})$ is the PDF

corresponding to a different process. Suppose $f(x_{[1:T]})$ is some function of the path, such as

$$f(x_{[1:T]}) = \max_{1\le k\le T} x_k \ .$$

The reweighting formula (3) with (2) applies in this case, with x replaced by $x_{[1:T]}$.

We calculate $L(x_{[1:T]})$ for the case when v corresponds to the symmetric random walk with no tendency to return to the origin:

$$X_{k+1} = X_k + bZ_k \ . \tag{10}$$

We need expressions for $u(x_{[1:T]})$ and $v(x_{[1:T]})$. We start with u, but the reasoning for v is similar. X_k being a Markov process we have the product formula for the joint density

$$u(x_1, \ldots, x_T) = u(x_1|x_0)\, u(x_2|x_1) \cdots u(x_T|x_{T-1}) \ .$$

Note that x_0 appears on the right but not on the left. That is because it is not a random variable. The transition densities $u(x_{k+1}|x_k)$ are Gaussian: $u(\cdot|x_k) = \mathcal{N}(ax_k, b^2)$. This leads to the formula

$$u(x_{k+1}|x_k) = \frac{1}{\sqrt{2\pi b^2}} e^{-(x_{k+1}-ax_k)^2/2b^2} \ .$$

Therefore,

$$u(x_{[1:T]}) = (2\pi)^{-T/2}\ b^{-T}\ \exp\left[\frac{-1}{2b^2}\sum_{k=0}^{T-1}(x_{k+1}-ax_k)^2\right] \ . \tag{11}$$

The expression for $v(x_{[1:T]})$ is similar, with $a = 1$.

Direct computation of $L(x_{[1:T]}) = u(x_{[1:T]})/v(x_{[1:T]})$ gives a starting point

$$L(x_{[1:T]}) = \exp\left\{\frac{-1}{2b^2}\sum_{k=0}^{T-1}\left[(x_{k+1}-ax_k)^2 - (x_{k+1}-x_k)^2\right]\right\} \ .$$

We do some algebra in the exponent

$$\begin{aligned}
&(x_{k+1}-ax_k)^2 - (x_{k+1}-x_k)^2 \\
&\qquad = (x_{k+1}-x_k+(1-a)x_k)^2 - (x_{k+1}-x_k)^2 \\
&\qquad = (x_{k+1}-x_k+(1-a)x_k)^2 - (x_{k+1}-x_k)^2 \\
&\qquad = 2(1-a)x_k\,(x_{k+1}-x_k) + (1-a)^2x_k^2 \ .
\end{aligned}$$

This leads to the expression for L we want:

$$L(x_{[1:T]}) = \exp\left[\frac{a-1}{b^2}\sum_{k=0}^{L-1} x_k\,(x_{k+1}-x_k)\right]\exp\left[\frac{-(a-1)^2}{2b^2}\sum_{k=0}^{L-1}x_k^2\right] \ . \tag{12}$$

Here are two Monte Carlo approaches to computing $\mathrm{E}\big[\, f(X_{[1:T]})\big]$. This is approximate C++ style code, with `norm()` producing a new standard normal each call. Figure 1 is the vanilla Monte Carlo with the process (9). Figure 1.1.2 is the vanilla Monte Carlo with the process (10) and likelihood ratio given by (12). Both codes (barring bugs) are "correct" in the sense that they converge to the exact answer as $N \to \infty$. But the importance sampler could be much more (or, alas, less) accurate for moderate N.

```
double X[T+1];           // a path, with X[0] given
X[0] = given value
sum = 0.;                // initialize
for j = 1, ... N {       // N = number of MC samples
  for k = 0, ...  T-1 {  // T = number of steps in a path
    X[k+1] = a*X[k] + b*norm();
   }
  sum = sum + f(X);      //  A function of the whole path
 }
A = sum/N;               //  The estimate of E[f(X)]
```

Figure 1: Vanilla Monte Carlo estimator of $\mathrm{E}[\, f(X)]$.

In finance, people often talk about different probability measures as this or that *world*. There is one world governed by (9) in which the process X_k is mean reverting. There is another world governed by (10) in which the X_k are a Gaussian martingale. The expected value of $f(X)$ in the mean reverting world is equal to the weighted expected value in the martingale world. The weight factor is (12).

1.1.3 Continuous time Gaussian process

The Ornstein Uhlenbeck process is

$$dX_t = -\gamma X_t dt + \sigma dW_t \; . \tag{13}$$

We want to express expectations of path functionals in this mean reverting world in terms of expectations in the world in which X_t is a Brownian motion. The problem is that continuous time processes do not have associated probability densities. It turns out that the likelihood ratio $L(x)$ can exist even the densities $u(x)$ and $v(x)$ do not.

A discrete approximation of the Ornstein Uhlenbeck process shows how this is possible. Use the usual conventions, small Δt, discrete times $t_k = k\Delta t$, and $X_k \approx X_{t_k}$. The approximations are $dX_t \leftarrow (X_{t_{k+1}} - X_{t_k}) \leftarrow (X_{k+1} - X_k)$, and $X_t dt \leftarrow X_k \Delta t$, and $dW_t \leftarrow \sqrt{\Delta t} Z_k$. We write the discrete approximation, then rewrite it in the form (9) to identify the parameters a and b as they depend on

```
double X[T+1];           // a path, with X[0] given
X[0] = given value
sum = 0.;                // initialize
for j = 1, ... N {       // N = number of MC samples

                                 // create a path
  for k = 0, ...  T-1 {          // T = number of steps in a path
    X[k+1] = X[k] + b*norm();    // The recurrence with a = 1
   }
                         //  calculate the likelihood function
  double s1 = 0.;        //  The sums in the likelihood fn.
  double s2 = 0.;
  for k = 0, ...  T-1 {  // T = number of steps in a path
    s1 = s1 + x[k]*(x[k+1] - x[k];
    s2 = s2 + x[k]*x[k];
   }
  L = exp( (a-1)/(b*b)*s1) * exp( (a-1)*(a-1)/(2*b*b)*s2 );
  sum = sum + L*f(X);    //  The function weighted by L(X)
 }
A = sum/N;               //  The estimate of E[f(X)]
```

Figure 2: Importance sampling Monte Carlo estimator of $\mathrm{E}_u[\,f(X)] = \mathrm{E}_v[\,L(X)f(X)]$.

Δt, γ, and σ:

$$\begin{aligned} X_{k+1} - X_k &= -\gamma X_k \Delta t + \sigma\sqrt{\Delta t} Z_k \\ X_{k+1} &= (1 - \gamma\Delta t)\, X_k + \sigma\sqrt{\Delta t} Z_k \end{aligned} \tag{14}$$

This has the form (9) with $a = (1 - \gamma\Delta t)$ and $b = \sigma\sqrt{\Delta t}$. The first exponent of (12) is

$$\frac{a-1}{b^2} \sum_{k=0}^{L-1} x_k \,(x_{k+1} - x_k) = \frac{-\gamma}{\sigma^2} \sum_{t_k < T} x_k \,(x_{k+1} - x_k) \to \frac{-\gamma}{\sigma^2} \int_0^T x_t \, dx_t \,. \tag{15}$$

The second exponent is

$$\frac{-(a-1)^2}{2b^2} \sum_{k=0}^{L-1} x_k^2 = \frac{-\gamma^2 \Delta t}{2\sigma^2} \sum_{k_k < T} x_k^2 \to \frac{-\gamma^2}{2\sigma^2} \int_0^T x_t^2 \, dt \,.$$

The limit as $\Delta t \to 0$ of the likelihood ratio seems to be

$$L(x_{[0,T]}) = e^{\frac{-\gamma}{\sigma^2} \int_0^T x_t \, dx_t} \, e^{\frac{-\gamma^2}{2\sigma^2} \int_0^T x_t^2 \, dt} \,. \tag{16}$$

That's a stochastic integral in the first exponent and a Riemann integral in the second.

It's common to denote measures in path space with capital letters such as P and Q. So, suppose the P world is one in which X_t is the Ornstein Uhlenbeck process (13), and Q is the world in which X_t is a Brownian motion with variance $\sigma^2 t$ and $X_0 = x_0$. Suppose $f(x_{[0,T]})$ is a path functional. The reweighting formula is

$$\mathrm{E}_P\big[\, f(X_{[0,T]})\big] = \mathrm{E}_Q\left[e^{\frac{-\gamma}{\sigma^2}\int_0^T X_t\, dX_t}\, e^{\frac{-\gamma^2}{2\sigma^2}\int_0^T X_t^2\, dt}\, f(X_{[0,T]})\right] \ . \tag{17}$$

1.2 Absolutely continuous, completely singular, ...

The reweighting formula for the Ornstein Uhlenbeck process is an example of something that happens often in fancy probability. You have two probability measures, P and Q, neither of which have densities, but which are related to each other through reweighting. The random outcome in the general abstract setting is $\omega \in \Omega$. Most of our examples have a path $x_{[0,T]}$ in place of ω as the random outcome. Suppose that $L(\omega)$ is a non-negative function so that for "every" function,

$$\mathrm{E}_P[\, f(\omega)] = \mathrm{E}_Q[\, f(\omega)L(\omega)] \ . \tag{18}$$

Integration with respect to a probability measure is written

$$\mathrm{E}_P[\, f(\omega)] = \int_\Omega f(\omega)\, dP(\omega) \ .$$

Therefore, we can express (18) as

$$\int_\Omega f(\omega)\, dP(\omega) = \int_\Omega f(\omega)L(\omega)\, dQ(\omega) \ . \tag{19}$$

Assuming we can put in "any" function f, this reduces to an identity about measures

$$dP(\omega) = L(\omega)dQ(\omega) \ ,$$

which is more commonly written

$$L(\omega) = \frac{dP(\omega)}{dQ(\omega)} \ . \tag{20}$$

The quantity on the right is called the *Radon Nikodym derivative* of measure P with respect to measure Q. If the random outcome is $x \in \mathbb{R}^n$ and the measures are given by probability densities $dP(x) = u(x)dx$ and $dQ(x) = v(x)dx$, then the Radon Nikodym derivative is the likelihood ratio

$$\frac{dP(x)}{dQ(x)} = \frac{u(x)dx}{v(x)dx} = \frac{u(x)}{v(x)} \ .$$

In abstract spaces, it's common that the left side is well defined even though there is no u and v. Of course, the actual definition of the formal ratio (20) is either the expectation relation (18), or the equivalent integral expression (19).

If there is such an L, we say that the measure P is *absolutely continuous* with respect to the measure Q.

For a given pair of probability measures P and Q, there may or may not be a function L relating them. Here's a simple test. Suppose there is an event, $A \subset \Omega$ with $P(A) = 1$ and $Q(A) = 0$. Choose $f(\omega) = \mathbf{1}_A(\omega)$. Clearly, $\mathbf{1}_A(\omega) \geq 0$ for all ω, and in the Q measure $\mathbf{1}_A(\omega) = 0$ almost surely. If $f(\omega) \geq 0$ for all ω, and $\mathrm{E}_Q[\, f(\omega)] = 0$, then $f(\omega) = 0$ almost surely with respect to Q. In particular, if $L(\omega)$ is any function, then $\mathbf{1}_A(\omega)L(\omega) = 0$ almost surely with respect to Q. This implies that

$$\mathrm{E}_P[\, \mathbf{1}_A(\omega)] = \mathrm{E}_Q[\, L(\omega)\mathbf{1}_A(\omega)] \ .$$

But the left side is one and the right side is zero. Therefore, there is no L in this case.

1.3 Girsanov's theorem

Suppose P and Q are measures in continuous time path space given by diffusion processes that satisfy

$$\left.\begin{array}{ll} P: & \mathrm{E}[\, \Delta X_t \mid \mathcal{F}_t] = a_P(X_t)\Delta t + \mu_P(X_t)\Delta t + o(\Delta t) \\ Q: & \mathrm{E}[\, \Delta X_t \mid \mathcal{F}_t] = a_Q(X_t)\Delta t + \mu_Q(X_t)\Delta t + o(\Delta t) \ . \end{array}\right\} \tag{21}$$

Girsanov's theorem determines when there is an $L(X_{[0,T]})$ so that for "any" path functional $f(x_{[0,T]})$,

$$\mathrm{E}_P\big[\, f(X_{[0,T]})\big] = \mathrm{E}_Q\big[\, L(X_{[0,T]})f(X_{[0,T]})\big] \ .$$

The answer is simple. If $\mu_P(x) > 0$ for all x, then P and Q are equivalent measures if and only if

$$\mu_P(x) = \mu_Q(x) \quad \text{for all } x. \tag{22}$$

As an "if and only if" theorem, Girsanov's theorem has two parts. The "only if" part is that if $\mu_P \neq \mu_Q$, then P and Q are completely singular with respect to each other. In other words, the equality of the noise coefficients is a *necessary condition* for the measures to be equivalent. We give the straightforward and quick argument for this first. The "if" part states that the equality (22) is a *sufficient condition* for P and Q to be equivalent. That is deeper and harder and, in practice, more interesting. The main point of the sufficiency argument is *Girsanov's formula*, which is an explicit stochastic integral formula for $L(x_{[0,T]})$.

The idea of the "only if" part appeared already in Section 1.2; you can recover the noise coefficient from a single path. For a general $\mu(x)$, this comes from the quadratic variation

$$[X]_t = \lim_{\Delta t \to 0} \sum_{t_k < t} \big(X_{t_{k+1}} - X_{t_k}\big)^2 \ , \tag{23}$$

and the formula

$$[X]_t = \int_0^t \mu(X_s)\, ds \ . \tag{24}$$

Now, let A be the set of paths $x_{[0,T]}$, such that the limit (23) exists, and satisfies

$$\frac{d}{dt}\,[X]_t = \mu_P(X_t)\ ,$$

for all t. This set has probability $P(A) = 1$ in the P process and probability $Q(A) = 0$ in the Q process. This shows that P and Q are completely singular with respect to each other.

To be very very picky, we could wonder what happens if $\mu_P(x) = \mu_Q(x)$ for some X values but not others. Consider the event

$$B = \{\mu_P(X_t) = \mu_Q(X_t) \text{ for all } t \in [0,T]\}\ .$$

You could imagine that this the probability of B is larger than zero but less than one. This is very very unlikely in a practical application. But if it would happen, the probability measures P and Q would not be equivalent, but would not be completely singular either. If $\mu_P(x) > c > 0$, then a P measure path has positive probability to reach any point, including points where $\mu_P \neq \mu_Q$.

1.3.1 Girsanov's formua, an informal derivation

There is an informal derivation of Girsanov's formula that is similar to the derivation of the Ornstein Uhlenbeck formula (16). We suppose that $\mu_P(x) = \mu_Q(x) = \mu(x)$ for all x and write an approximate formulas that correspond to the exact ones for Gaussian processes, (12) and (15), etc. We put ourselves in the P world and assume also that the transition densities for X_t in the P world are approximately Gaussian. Define $x^{\Delta t}_{[1:n_T]}$ as the discrete time "observations" of the continuous time proces x_t. The values are $x^{\Delta t}_k = x_{t_k}$, with $t_k = k\Delta t$. The number of observations is $n_T = \max\{t_k \text{ with } t_k < T\}$. We denote the PFD of $x^{\Delta t}_{[1:n_T]} \in \mathbb{R}^{n_T}$ in the P or Q measures as $U_P(x^{\Delta t}_{[1:n_T]})$ and $U_Q(x^{\Delta t}_{[1:n_T]})$ respectively. The observation sequence forms a Markov chain, so the joint PDF is the product of transition probabilities

$$u(x^{\Delta t}_{k+1} \mid x^{\Delta t}_k)\ .$$

The transition densities are approximately Gaussian if Δt is small. The conditional mean and variance is given by (21). The formulas that correspond to the data are

$$\begin{aligned} X^{\Delta t}_{k+1} &\sim \mathcal{N}(X^{\Delta t}_k + a_P(X^{\Delta t}_k)\Delta t, \mu(X^{\Delta t}_k)\Delta t) \qquad \text{(approximately)}, \\ u_P(x^{\Delta t}_{k+1} \mid x^{\Delta t}_k) &\approx \frac{1}{\sqrt{2\pi\mu(x^{\Delta t}_k)\Delta t)}} \exp\left\{ \frac{-\left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_P(x^{\Delta t}_k)\Delta t\right)^2}{2\mu(x^{\Delta t}_k)\Delta t} \right\}\ . \end{aligned} \tag{25}$$

We get

$$L^{\Delta t}(x^{\Delta t}_{[1:n_T]}) = \frac{U_P(x^{\Delta t}_{[1:n_T]})}{U_Q(x^{\Delta t}_{[1:n_T]})}$$

as we did in Section 1.1.3 by multiplying the Gaussian factors (25) and cancelling common terms. The prefactors cancel from the numerator and denominator because $\mu_p = \mu_Q$. We write the result as

$$L^{\Delta t}(x^{\Delta t}_{[1:n_T]}) = e^{H(x^{\Delta t}_{[1:n_T]})} \ ,$$

with an exponent given by

$$H(x^{\Delta t}_{[1:n_T]}) = \sum_{t_k<T} \frac{-\left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_P(x^{\Delta t}_k)\Delta t\right)^2 + \left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_Q(x^{\Delta t}_k)\Delta t\right)^2}{2\mu(x^{\Delta t}_k)\Delta t} \ .$$

The algebra continues as in Section 1.1.3 by simplifying the difference in the numerator

$$\begin{aligned}
&-\left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_P(x^{\Delta t}_k)\Delta t\right)^2 + \left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_Q(x^{\Delta t}_k)\Delta t\right)^2 \\
&\qquad = -\left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_Q(x^{\Delta t}_k)\Delta t - \left[a_P(x^{\Delta t}_k) - a_Q(x^{\Delta t}_k)\right]\Delta t\right)^2 \\
&\qquad\qquad\qquad + \left(x^{\Delta t}_{k+1} - x^{\Delta t}_k - a_Q(x^{\Delta t}_k)\Delta t\right)^2 \\
&\qquad = 2\left(x^{\Delta t}_{k+1} - x^{\Delta t}_k\right)\left[a_P(x^{\Delta t}_k) - a_Q(x^{\Delta t}_k)\right]\Delta t \\
&\qquad\quad - \left[a_P(x^{\Delta t}_k) - a_Q(x^{\Delta t}_k)\right] a_Q(x^{\Delta t}_k)\Delta t^2 \ .
\end{aligned}$$

We plug these two last terms into the H expression and get H as a sum of terms that correspond to the stochastic integral and Riemann integral terms in Section 1.1.3; $H(x^{\Delta t}_{[1:n_T]}) = H_S(x^{\Delta t}_{[1:n_T]}) + H_R(x^{\Delta t}_{[1:n_T]})$, where

$$H_S(x^{\Delta t}_{[1:n_T]}) = \sum_{t_k<T} \frac{\left(x^{\Delta t}_{k+1} - x^{\Delta t}_k\right)\left[a_P(x^{\Delta t}_k) - a_Q(x^{\Delta t}_k)\right]}{\mu(x^{\Delta t}_k)} \ , \tag{26}$$

and

$$H_R(x^{\Delta t}_{[1:n_T]}) = \sum_{t_k<T} \frac{-\left[a_P(x^{\Delta t}_k) - a_Q(x^{\Delta t}_k)\right] a_Q(x^{\Delta t}_k)}{2\mu(x^{\Delta t}_k)}\,\Delta t \ . \tag{27}$$

The $\Delta t \to 0$ limits of these are

$$H_S(x_{[0,T]}) = \int_0^T \frac{a_P(x_t) - a_Q(x_t)}{\mu(x_t)}\,dx_t \ ,$$

and

$$H_R(x_{[0,T]}) = -\int_0^T \frac{\left[a_P(x_t) - a_Q(x_t)\right] a_Q(x_t)}{2\mu(x_t)}\,dt \ .$$

Finally, we substitute these to get the $\Delta t \to 0$ limit of $L^{\Delta t}$. The result is Girsanov's formula

$$L(x_{[0,t]}) = \exp\left(\int_0^t \frac{a_P(x_s) - a_Q(x_s)}{\mu(x_s)}\,dx_s\right)\exp\left(-\int_0^t \frac{\left[a_P(x_s) - a_Q(x_s)\right] a_Q(x_s)}{2\mu(x_s)}\,ds\right) \ . \tag{28}$$

1.3.2 Digression on Ito differentiation

We will be doing some Ito calculations in the next sections. It is useful to state the general Ito differentiation rule that are needed. There is really just one differentiation rule, but a multi-component one. This single version of Ito's lemma leads to specific rules for differentiating products, quotients, exponentials, etc.

There are many differentiation rules in ordinary calculus too. But with multivariate calculus you can get by with just one, the chain rule for partial derivatives. If $g(x) = f(u(x), v(x))$, then

$$\frac{dg}{dx} = \partial_u f(u(x), v(x)) \frac{du}{dx} + \partial_v f(u(x), v(x)) \frac{dv}{dx} \ . \tag{29}$$

The partial derivatives on the right are called *partial* because each one determines a part of the total change in f. There are the parts coming from the change in u and the change in v respectively. According to the chain rule, the total change in f is the sum of the two partial changes, one that comes from changing u but fixing v, and one that comes from fixing u and changing v.

We illustrate by giving a complicated derivation of a simple differentiation formula. take $f(u, v) = uv$, and $u(x) = x$ and $v(x) = x$. Then $g(x) = u(x)v(x) = x^2$. The derivative formula (29) then gives

$$\frac{d}{dx} x^2 = \left(\frac{du}{dx}\right) x + x \left(\frac{dv}{dx}\right) = 1 \cdot x + x \cdot 1 = 2x \ .$$

The general version of Ito's lemma is as follows. Let $X_t \in \mathbb{R}^n$ be an $n-$component Ito process. Suppose $Y_t \in \mathbb{R}^m = f(X_t)$. Suppose that all partial derivatives up to second order of $f(x)$ make sense. Suppose that the infinitesimal drift and noise coefficients of X_t are

$$\begin{aligned} \mathrm{E}[\, \Delta X_t \mid \mathcal{F}_t] &= a_t^X \Delta t + o(\Delta t) \\ \mathrm{E}\Big[\, (\Delta X_t)\, (\Delta X_t)^t \mid \mathcal{F}_t \Big] &= \mu_t^X \Delta t + o(\Delta t) \ , \end{aligned}$$

where a_t^X is an $n-$component vector and μ_t^X is an $n \times n$ non-negative symmetric matrix. Then Y_t is an Ito process with drift and noise given by

$$\begin{aligned} \mathrm{E}[\, \Delta Y_t \mid \mathcal{F}_t] &= a_t^Y \Delta t + o(\Delta t) \\ \mathrm{E}\Big[\, (\Delta Y_t)\, (\Delta Y_t)^t \mid \mathcal{F}_t \Big] &= \mu_t^Y \Delta t + o(\Delta t) \ . \end{aligned}$$

We get a_t^Y and μ_t^Y as follows. Differentiate $Y_{j,t} = f_j(X_t)$ to get (using the summation convention for indices k and l)

$$dY_{j,t} = df_j(X_t) = \frac{\partial f_j}{\partial x_k} dX_{k,t} + \frac{1}{2} \frac{\partial^2 f_j}{\partial x_k \partial x_l} dX_{k,t} dX_{l,t} \ .$$

The expectation of $dY_{j,t}$ is

$$a_{j,t}^Y dt = \mathrm{E}[\, dY_{j,t} \mid \mathcal{F}_t] = \left(\frac{\partial f_j}{\partial x_k} a_{k,t}^X + \frac{1}{2} \frac{\partial^2 f_j}{\partial x_k \partial x_l} \mu_{kl,t}^X \right) dt \ .$$

The noise coefficient for Y is (summing over l and m)

$$\begin{aligned}\mu^Y_{jk,t}dt &= \mathrm{E}[\, dY_{j,t}dY_{k,t} \mid \mathcal{F}_t] \\ &= \frac{\partial f_j}{\partial x_l}\frac{\partial f_k}{\partial x_m}\mathrm{E}[\, dX_l dX_m \mid \mathcal{F}_t] \\ &= \frac{\partial f_j}{\partial x_l}\frac{\partial f_k}{\partial x_m}\mu^X_{lm,t}dt\ .\end{aligned}$$

You can put these formulas in matrix form, such as

$$\mu^Y_t = (Df(X_t))\,\mu^X_t\,(Df(X_t))^t\ .$$

Where Df is the Jacobian matrix of f. We will not use these formulas as much as the computations behind them in specific cases.

As a first application, we compute the differential of

$$Y_t = e^{\int_0^t m_s dW_s}\ .$$

The Ito process in the exponent is

$$X_t = \int_0^t m_s dW_s\ .$$

This satisfies $dX_t = m_t dW_t$. The "Ito term" involves $\mathrm{E}\Big[\,(dX_t)^2 \mid \mathcal{F}_t\Big] = m_t^2 dt$. The differential of Y_t is

$$\begin{aligned}dY_t &= \frac{\partial e^x}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 e^x}{\partial x^2}\mathrm{E}\Big[\,(dX_t)^2\Big] \\ &= e^X_t m_t dW_t + \frac{1}{2}e^{X_t}m_t^2 dt \\ d\left(e^{\int_0^t m_s dW_s}\right) &= e^{\int_0^t m_s dW_s}m_t dW_t + \frac{1}{2}e^{\int_0^t m_s dW_s}m_t^2 dt\ .\end{aligned}$$

Here's a more complicated example, the ratio of the value at time t to the average value

$$Y_t = \frac{1}{X_t}\,\frac{1}{t}\int_0^t X_s ds\ .$$

Suppose X_t is a geometric Brownian motion that satisfies $X_t = rX_t dt + \sigma X_t dW_t$. To compute the differential, define $U_t = X_t$ and $V_t = \frac{1}{t}\int_0^t X_s ds$. Define $f(u,v) = \frac{v}{u}$. For differentiation, we need

$$\begin{aligned}\partial_u f(u,v) &= -\frac{v}{u^2} \\ \partial_v f(u,v) &= \frac{1}{u} \\ \partial_u^2 f(u,v) &= \frac{v}{u^3} \\ \partial_u\partial_v f(u,v) &= -\frac{1}{u^2} \\ \partial_v^2 f(u,v) &= 0\ .\end{aligned}$$

The relevant differentials are

$$\begin{aligned} dU_t &= dX_t = rX_t dt + \sigma X_t dW_t \\ dV_t &= \left(\frac{-1}{t^2} \int_0^t X_s ds \right) dt + \frac{1}{t} X_t dt \\ \mathrm{E}\Big[\, (dU_t)^2 \mid \mathcal{F}_t \Big] &= \sigma^2 X_t^2 dt \\ \mathrm{E}[\, dU_t dV_t \mid \mathcal{F}_t] &= 0 \\ \mathrm{E}\Big[\, (dV_t)^2 \mid \mathcal{F}_t \Big] &= 0 \; . \end{aligned}$$

Therefore,

$$\begin{aligned} dY_t = & \frac{-rX_t dt - \sigma X_t dW_t}{tX_t^2} \int_0^t X_s ds \\ & + \frac{-1}{t^2 X_t} \left(\int_0^t X_s ds \right) dt + \frac{dt}{t} + \frac{\sigma^2}{tX_t} \left(\int_0^t X_s ds \right) dt \; . \end{aligned}$$

This is not a beautiful formula. It just shows the workings of the Ito calculus in a systematic if uninspired way. If the integral in the numerator had been

$$V_t = \int_0^t X_s dW_s \; ,$$

then we would have had $\mathrm{E}[\, dU_t dV_t \mid \mathcal{F}_t] = \sigma X_t^2 dt$.

1.3.3 Digression on weak and strong solutions

There are two reasons we might write

$$dX_t = b(X_t) dW_t \; . \tag{30}$$

The more obvious reason would be that X_t is a function of $W_{[0,t]}$, and the differential of X_t is $b(X_t)dW_t$. The less obvious and less correct but more common reason is to express the view that X_t is a diffusion process that satisfies

$$\mathrm{E}[\, dX_t \mid \mathcal{F}_t] = 0 \;\; , \quad \mathrm{E}\Big[\, (dX_t)^2 \mid \mathcal{F}_t \Big] = b(X_t)^2 dt \; . \tag{31}$$

The first interpretation is called *strong*. The second is *weak*. It comes up in stochastic modeling – we think we know the infinitesimal mean and variance of a stochastic process. There is no hypothetical Brownian motion W_t in the model, only a hypothetical infinitesimal mean and variance (31). The strong form (30) is just a convenient way of expressing the weak form (31). The strong form is more convenient largely because it's shorter and seems less technical.

There is a mathematical connection between the forms that gives an excuse, sort of, for confusing weak and strong forms. This is a *martingale representation theorem*, which goes something like this. Suppose there is a filtration $\mathcal{F}_t$ and

an Ito process that is a martingale and satisfies (31). Then there is a Brownian motion W_t that is progressively measurable with respect to $\mathcal{F}_t$ so that

$$X_t - X_0 = \int_0^t b_s dW_s \ ,$$

where b_s is the square root of the b_s^2 in (31). Most of our theory – in particular the generator, forward and backward equations, etc., require only the weak form.

The martingale representation theorem is not much harder than the *Levi uniqueness theorem*: if X_t is an Ito process with $\mathrm{E}\Big[\, (dX_t)^2 \mid \mathcal{F}_t \Big] = dt$, then X_t is a standard Brownian motion with Gaussian transition densities and all the other properties of Brownian motion. This is sort of a central limit theorem. We know the conditional variance is $\mathrm{E}\Big[\, (X_t - X_s)^2 \mid \mathcal{F}_s \Big] = (t-s)$. This follows from the infinitesimal variance property. We also know that increments of X over disjoint intervals are uncorrelated. But we don't know these increments are independent, or Gaussian. Levi has a clever proof that is like one of the proofs of the ordinary CLT.

1.3.4 Application to Brownian motion

Suppose the Q measure is Brownian motion and the P measure is Brownian motion with a fixed drift velocity a. In the general formula (28), we have $\mu = 1$, and $a_Q = 0$, and $a_P = a$. The first integral in (28) simplifies to

$$a \int_0^t dx_s = a(x_t - x_0) \ .$$

The second integral simplifies to $-a^2 t/2$. Assuming $x_0 = 0$, this leads to

$$\mathrm{E}_a[\, f(X_t)] = e^{-a^2 t/2} \mathrm{E}_{\mathrm{BM}}\big[\, f(X_t) e^{aX_t}\big] \ .$$

Let $u_P(x,t)$ and $u_Q(x,t)$ be the probability densities for X_t under Brownian motion with drift (P) and without drift (Q). We then have

$$u_P(x,t) = e^{-a^2 t/2} e^{ax} u_Q(x,t) \ . \tag{32}$$

We have seen this relationship already. The forward PDE for u_P is

$$\partial_t u_P = L_P^* u_P = \frac{1}{2}\partial_x^2 u_P - a \partial_x u_P \ , \tag{33}$$

while the forward equation for u_Q is just the heat equation

$$\partial_t u_Q = L_Q^* u_Q = \frac{1}{2}\partial_x^2 u_Q \ . \tag{34}$$

The change of variables (32) converts a solution of (34) to a solution of (33).

1.3.5 Reweighting of Brownian motion

There is an approach to Girsanov re-weighting that does the hard part only for Brownian motion, a case where the hard part is easier. Suppose that in the Q measure W_t is a standard Brownian motion. We seek a re-weighting to give W_t a net drift with rate $m_t = m(W_t, t)$. That is, in the P measure,

$$\mathrm{E}_P[\, dW_t \mid \mathcal{F}_t] = m(W_t, t)dt = m_t dt \; . \tag{35}$$

The change of measure formula (28) simplifies in this case. It leads you to consider the integral

$$L_t = \exp\left[\int_0^t m_s dW_s - \frac{1}{2} \int_0^t m_s^2 ds \right] \; . \tag{36}$$

We can verify that this formula accomplishes (35). It pulls the mean of dW_t from 0 to $m_t dt$. To verify the sign, suppose $m_t > 0$, then the integral (36) gives paths W_t the go up more weight than paths that go down.

We write $m_t = m(W_t, t)$ so the re-weighted process will be a Markov process. The formulas are more complicated if the drift at time t is a function of the path at earlier times.

The first step of the verification is that $\mathrm{E}_Q[\, L_t] = 1$. This is necessary because $\mathrm{E}_P[\, 1] = \mathrm{E}_Q[\, L_t]$. We do this by showing that L_t is a martingale with respect to Q. Since $L_0 = 1$, this implies that $\mathrm{E}_Q[\, L_t] = 1$. This is an exercise in our general Ito calculus. The exponent is

$$H_t = \int_0^t m_s dW_s - \frac{1}{2} \int_0^t m_s^2 ds \; .$$

Its differential is $dH_t = m_t dW_t - \frac{1}{2} m_t^2 dt$. The differential of $L_t = e^{H_t}$ is

$$dL_t = e^{H_t} dH_t + \frac{1}{2} e^{H_t} \mathrm{E}\Big[\, (dH_t)^2 \mid \mathcal{F}_t \Big] \; .$$

It does not matter whether the expectation $\mathrm{E}\big[\, dH_t^2 \mid \mathcal{F}_t \big]$ is taken with respect to P or Q, because re-weighting changes only $\mathrm{E}[\, dH_t \mid \mathcal{F}_t]$, not $\mathrm{E}\Big[\, (dH_t)^2 \mid \mathcal{F}_t \Big]$. We have $\mathrm{E}\Big[\, (dH_t)^2 \mid \mathcal{F}_t \Big] = m_t^2 dt$ in either measure. These dH_t calculations can be substituted to find $dL_t = m_t dW_t$, which implies that L_t is a martingale.

The next calculation verifies (35). This is a conditional expectation. We need to verify that if you condition on $\mathcal{F}_t$, the change of measure formula is the natural. A Markov process "starts over" at each time t, so the weighting should have the same property. If $F(w_{[t_1,t_2]})$ is a function that depends on the path only between times t_1 and t_2, then

$$\mathrm{E}_P\big[\, F(W_{[t_1,t_2]}) \mid \mathcal{F}_{t_1} \big] = \mathrm{E}_Q\left[F(W_{[t_1,t_2]}) \exp\left(\int_{t_1}^{t_2} m_t dW_t - \int_{t_1}^{t_2} \frac{1}{2} m_t^2 dt \right) \mid \mathcal{F}_{t_1} \right] \; .$$

This formula is "obvious". It is easy to check using the simple formula for conditional expectation in discrete cases. We may have more to say about it below.

Assuming the formula, the calculation is

$$\begin{aligned}
\mathrm{E}_P[\, dW_t \mid \mathcal{F}_t] &= \mathrm{E}_Q\left[\, dW_t e^{\int_t^{t+dt} m_t dW_t - \frac{1}{2}\int_t^{t+dt} m_t^2 dt} \mid \mathcal{F}_t \right] \\
&= \mathrm{E}_Q\left[\, dW_t \left(1 + m_t dW_t - -\frac{1}{2} m_t^2 dt \right) \mid \mathcal{F}_t \right] \\
&= \mathrm{E}_Q[\, dW_t \mid \mathcal{F}_t] + m_t \mathrm{E}_Q\left[\, dW_t^2 \mid \mathcal{F}_t \right] \\
&= m_t dt \ .
\end{aligned}$$

The general change of measure follows from this as follows. In the Q measure, W_t is a Brownian motion. But in the P measure, W_t has a drift $m_t dt$. A process with that drift removed is

$$dB_t = dW_t - m_t dt \ . \tag{37}$$

In the P measure, B_t is a standard Brownian motion. Now suppose a process is described in the P measure by

$$dX_t = a_P(X_t)dt + b_P(X_t)dB_t \ .$$

This means that in the P "world", X_t is a diffusion process with drift $a_P(x)dt$ and infinitesimal variance $b_P(x)^2 dt$. We find the description of X_t in the Q measure by substituting (37)

$$\begin{aligned}
dX_t &= a_P(X_t)dt + b_P(X_t)\,[dW_t - m_t dt] \\
&= [a_P(X_t) - m_t b_P(X_t)]\, dt + b_P(X_t) dW_t \ .
\end{aligned}$$

In the Q measure, dW_t is a standard Brownian motion. this implies that in the Q measure X_t has drift $a_Q = a_P(X_t) - m_t b_P(X_t)$. This again shows that you can adjust the drift but cannot change the infinitesimal variance.

Reweighting preserves sets of probability zero. If A is an event, and if $L(x) = \frac{dP(x)}{dQ(x)}$ is a reweighting formula that turns Q into P, then $P(A) = 0$ if and only if $Q(A) = 0$. For example, suppose $X_0 =$ and A is the event that $X_t > 0$ for all $t < \varepsilon$. More precisely, A is the event that there is an $\varepsilon > 0$ so that $X_t > 0$ for all $t < \varepsilon$. It is easy to show $Q(A) = 0$ if Q is Brownian motion measure. Therefore, $P(A) = 0$ for any measure equivalent to Brownian motion measure, which includes fixed and time dependent drift. Suppose b_t is a function of t with $|\frac{d}{dt} b_t| < \infty$ and $b_0 = 0$. Let A be the event $X_t > b_t$ for all $t < \varepsilon$. Define a new process $Y_t = X_t - b_t$. If the X process is equivalent to Brownian motion measure, then the Y process is too (Girsanov). In that case, $P(A) = 0$ here too.