Week 6

Jonathan Goodman, November, 2020

Tentative (more text, no more exercises)

The two topics for this week both involve putting integrals in the exponential. For the *Feynman Kac formula*, the integral is dt. For the *Girsanov change of measure formula*, the integral is dW_t. If X_t is an Ito process, then another Ito process is

$$Y_t = e^{X_t} \ .$$

We can apply Ito's lemma with $f(x) = e^x$, and $\partial_x f(x) = e^x = f(x)$, and $\partial_x^2 f(x) = f(x)$. The result is, using $Y_t = e^{X_t}$ on the right side

$$dY_t = Y_t \, dX_t + \frac{1}{2} Y_t \, (dX_t)^2 \ . \tag{1}$$

If X_t is a dt integral (Feynman Kac), then the second term on the right is zero.

The Feynman Kac "formula" is used in finance to study models of fluctuating interest rates. Suppose the interest rate at time t is X_t, and a "money market account" at time t has value M_t. Then $M_{t+dt} = (1+X_t dt)M_t$. This is expressed as

$$M_T = M_0 e^{\int_0^T X_t dt} \ .$$

The Feynman Kac formula is related to expected values of expressions like this.

1 Feynman Kac Formula

The *Feynman Kac* formula, as the term is used in finance, is a relationship between a multiplicative value function and the backward equation it satisfies. Suppose X_t is a diffusion process that satisfies the SDE

$$dX_t = a(X_t) \, dt + b(X_t) \, dW_t \ . \tag{2}$$

A multiplicative functional is a function of the diffusion process path of the form

$$e^{\int_0^T V(X_s) \, ds} \ . \tag{3}$$

This is called *multiplicative* because you can think of it as multiplying together the individual pieces $V(X_s)ds$. You can see this more explicitly by choosing a finite Δt approximation to the integral. As usual, $t_k = k\Delta t$. We can approximate the functional as

$$
\begin{aligned}
e^{\int_0^T V(X_s)\,ds} &\approx \exp\left\{\sum_{t_k<T} V(X_{t_k})\Delta t\right\} \\
&= \prod_{t_k<T} e^{V(X_{t_k})\Delta t} \;. \qquad (4)
\end{aligned}
$$

Moreover, you can use the exponential approximation $e^\epsilon \approx 1+\epsilon$ to write the approximation is

$$
\prod_{t_k<T} e^{V(X_{t_k})\Delta t} \approx \prod_{t_k<T} \left[\,1 + V(X_{t_k})\,\Delta t\,\right] \;.
$$

To be even more explicit, suppose you define the product of the first n factors to be

$$
P_n = \prod_{k=0}^{n-1} \left[\,1 + V(X_{t_k})\,\Delta t\,\right] \;.
$$

Then

$$
P_{n+1} = \left[\,1 + V(X_{t_n})\,\Delta t\,\right] P_n \;.
$$

This has the interpretation as an interest payment. The "account" at time n is P_n. As we go from time n to time $n+1$, we receive a payment of $V_n \Delta t P_n$. This is the payment, for example, if V is an interest rate in years and Δt is measured as a fraction of a year. The final amount P_n is random because the interest rates $V_k = V(X_{t_k})$ are random.

The value function corresponding to the multiplicative functional is

$$
f(x,t) = \mathrm{E}\left[\, e^{\int_t^T V(X_s)\,ds} \mid X_t = t \,\right] \;. \qquad (5)
$$

There may be other ways to define a value function if we are interested in the functional with the integral starting at time $t=0$. If you take the integral starting at time t, then functional in the value function (5) does not depend on X_s for $s<t$. That means that the expectation on the right is a function of x and t alone, not the path that got you there.

We can find the backward equation PDE that f satisfies using a variant of the martingale method we used in Week 4. Consider a small time dt, take d of both sides, use Ito's lemma, and set the dt parts equal, ignoring the dW part. On the left, you have

$$
df(X_t,t) = \partial_t f dt + \partial_x f\, dX + \frac{1}{2}\partial_x^2 f (dX)^2 \;.
$$

We identify the infinitesimal mean (the dt part) from the SDE (2) using the Ito rule $(dX)^2 = b^2 dt$. The result is

$$
df(X_t,t) = \partial_t f dt + a(x)\partial_x f dt + \frac{1}{2} b^2(x) \partial_x^2 f + \partial_x f b(x) dW_t \;.
$$

On the right, we can use the Ito calclation (1) to get

$$
\begin{aligned}
de^{\int_t^T V(X_s)\,ds} &= d\left(\int_t^T V(X_s)ds\right) e^{\int_t^T V(X_s)\,ds} \\
&= -V(x)\,dt\,e^{\int_t^T V(X_s)\,ds} \; .
\end{aligned}
$$

For the second line, note that if $V>0$, the integral gets smaller as t increases, and that we are conditioning on $X_t = x$. We put this into the right side of (5), combine terms, drop the common dt factor, and get

$$
\partial_t f(x,t) + a(x)\partial_x f(x,t) + \frac{1}{2}b^2(x)\partial_x^2 f(x,t) + V(x)f(x,t) = 0 \; . \tag{6}
$$

This is the backward equation for the multiplicative functional. The final condition "obviously" (think about it) is $V(x,T)=1$.

The examples are all rather complicated, unfortunately. Take X_t to be Brownian motion and $V(x)=x$. Then the backward equation is

$$
\partial_t f + \frac{1}{2}\partial_x^2 f + xf = 0 \; . \tag{7}
$$

You can calculate the solution explicitly as in Exercise 2. Or you can guess by trial and error. Either way, we come to the ansatz

$$
f(x,t) = e^{A(t)x+B(t)} \; .
$$

We plug this into the backward equation (7) using the calculations

$$
\begin{aligned}
\partial_t f &= \left(\dot{A}x + \dot{B}\right) f \\
\partial_x^2 f &= \partial_x\left[Ae^{Ax+B}\right] = A^2 f \\
xf &= xf \; .
\end{aligned}
$$

This gives

$$
\dot{A}xf + \dot{B}f + \frac{1}{2}A^2 f + xf = 0 \; .
$$

This works if we set the coefficient of x to zero, and the constant term (as a function of x) to zero. This gives the equations

$$
\begin{aligned}
\dot{A} + 1 &= 0 \\
\dot{B} + \frac{1}{2}A^2 &= 0 \; .
\end{aligned}
$$

The first equation, together with the final condition $A(T)=0$ (why?)) gives $A(t) = T-t$. The second equation becomes

$$
\dot{B} = -\frac{1}{2}(T-t)^2 \; .
$$

The solution is $B(t) = \frac{1}{6}(T-t)^3$. The full solution is

$$f(x,t) = e^{\frac{1}{6}(T-t)^3} e^{(T-t)x} \ .$$

The history of this formula starts in the late 1940's when physicist Richard Feynman wrote a formula for the solution of the Schrödinger equation as a "path integral", which was a kind of infinite product of infinitesimal pieces something like (4). Mathematicians reacted badly to this "Feynman integral" because it does not correspond to an integral in the sense of measure theory. About 15 years later, mathematician Mark Kac found that a similar formula gave an expression for the solution of the backward equation (6). The part of Feynman's formula that mathematicians objected to turned into "Wiener measure", which is the probability distribution of Brownian motion paths. This had the advantage of being mathematically rigorous. By the way, you pronounce "Kac" as "cats". His name is Polish. People who immigrated from Poland before him spelled their names as "Katz". For example, there is a "Katz Delicatessen" on Houston Street not far from NYU.

2 Interest rate models

The financial services industry (as opposed to pure traders) is heavily focused on loans and financial instruments that depend on interest rates. There is a range of stochastic models of interest rates, but the simple models involve only the *short rate* (also called *overnight rate*), which is the interest rate (in %/year, say) that is paid on a loan where you "know" it will be repaid very soon (e.g., the next day). Let us call this rate R_t. One model is

$$dR_t = -\gamma(R_t - r_0)\, dt + \sigma dW_t \ . \tag{8}$$

This is a mean-reverting process centered on an average long term average rate r_0.

A *floating rate* loan is a loan so that the interest rate at time t is related to an index of short rate loans at that time such as LIBOR (google this). We could use the model (8) as a model of this fluctuating rate. Suppose you make a loan with floating interest rate R_t to be repaid at time T. The total repayment amount is a multiplicative function of the initial amount

$$M_T = M_0 e^{\int_0^T R_s\, ds} \ .$$

We may as well set $M_0 = 1$ for simplicity. The value function satisfies the backward equation is

$$\partial_t f - \gamma(r - r_0)\partial_r f + \frac{\sigma^2}{2}\partial_r^2 f - rf = 0 \ .$$

This has an ansatz solution of the form

$$f(r,t) = e^{A(t)r + B(t)} \ .$$

Try it and see. Models like this are called *affine* because the exponent is an affine function of r.

3 Change of measure

Change of measure means representing an expected value with respect to one probability distribution with the expected value with respect to a different distribution. To get the idea, suppose $p(x)$ and $q(x)$ are two probability density functions for a one component random variable X. The expected value of $V(X)$ with respect to the p distribution is

$$\mathrm{E}_p[\, V(X)] = \int V(x)\, p(x) \; .$$

The *likelihood ratio* between p and q is

$$L(x) = \frac{p(x)}{q(x)} \; .$$

For now we assume that $q(x) \neq 0$ if $p(x) \neq 0$ so $L(x)$ doesn't have "denominator difficulties". The p expectation may be written as

$$\int V(x)\, p(x) = \int V(x)\, \frac{p(x)}{q(x)}\, q(x)\, dx \; .$$

This leads to the *change of measure formula*

$$\mathrm{E}_p[\, V(X)] = \mathrm{E}_q[\, V(X) L(X)] \; . \tag{9}$$

Finance people sometimes talk about "worlds". There can be a theory like the Black Scholes theory that represents a quantity F in terms of a random variable. For example (Week 5), the Black Scholes theory represents an option price as an expectation. If $F = \mathrm{E}_p[\, V(X)]$, they would say: "In the $p-$world F is the expected value of $V(X)$, but in the $q-$world, F is the expected value of $V(X)L(X)$. Probability distributions are called *measures*, so we talk about the $p-$measure and the $q-$measure.

As an example, suppose the $p-$measure says that $X \sim \mathcal{N}(0,1)$ and we want the $p-$world probability that $X > a$. This may be evaluated using the $q-$measure $X \sim \mathcal{N}(\mu,1)$. The probability densities are

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \;\; , \quad q(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} \; .$$

The ratio is

$$\begin{aligned} L(x) &= \frac{p(x)}{q(x)} \\ &= \exp\!\left(\frac{1}{2} \left[(x-\mu)^2 - x^2 \right] \right) \\ L(x) &= e^{-\mu x + \frac{1}{2}\mu^2} \; . \end{aligned}$$

The probability of $X > a$ involves the indicator function $V(x) = 1$ if $x > a$ and $V(x) = 0$ otherwise. Then

$$\Pr_p(X > a) = \int V(x)\, p(x)\, dx = \int V(x) L(x) q(x)\, dx$$

The expected value in terms of the $q-$measure is given by

$$\Pr_p(X > a) = e^{\frac{1}{2}\mu}\, \mathrm{E}_q\big[\, e^{-\mu X} V(X)\big] \ . \tag{10}$$

The extra complexity of the left side of (10) can give a more accurate way to calculate the probability than the direct approach suggested by the right side. The $p-$world variance of $V(X)$ can be much higher than the $q-$world variance of $e^{\frac{1}{2}\mu} e^{\mu X} V(X)$. You can estimate the probability on the left side by taking n independent samples $X_k \sim \mathcal{N}(0,1)$ and counting the number of "hits" (samples with $X_k > a$). The variance of this is $\frac{1}{n}p(1-p) \approx \frac{1}{n}p$ (for p small, which is the interesting case). The variance of the right side can be much smaller, as the calculations of Exercise 1 show.

3.1 Almost surely

This subsection is a sequence of examples to motivate Subsection 3.3. Suppose A is an event determined by some random variables. We say that A happens *almost surely* if

$$\Pr(A) = 1 \ .$$

For example, if $X \sim \mathcal{N}(0,1)$, then $\Pr(X \geq 0) = 1$. This is a silly example, because it has nothing to do with probability and because you don't need to say "almost". Any number x has $x \geq 0$, surely.

The strict inequality $X > 0$ is another event that happens almost surely. This one is less trivial, since $X = 0$ is possible in the sense that it is one of the outcomes in the probability model. However, $X = 0$ "never happens" in the sense that

$$\Pr(X = 0, \ \text{exactly}\ ) = 0 \ .$$

Unlike the trivial example $X \geq 0$, this one depends on the probability distribution of X. For example, suppose you "toss a coin" and take $X = 0$ with probability $\frac{1}{2}$ and $X \sim \mathcal{N}(0,1)$ with probability $\frac{1}{2}$.

$$\begin{aligned} Z &\sim \mathcal{N}(0,1) \\ U &= \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases} \\ X &= UZ \ . \end{aligned}$$

Such a random variable could come up in a financial model. We could imagine that a market is open with probability $\frac{1}{2}$ and if it is open, the price of an asset could change.

Harder examples come from probability distributions defined by limits (like Brownian motion) or distributions that depend on infinitely many variables. As an example of an infinitely-many-variables distribution, consider a sequence of coin tosses $U_1, U_2, \cdots, U_n, \cdots$. Suppose that $U_k = 1$ with probability p and $U_k = 0$ with probability $1-p$, and all the U_k are independent. You can ask: what is the probability that $U_k = 1$ for all k. The answer is $\Pr(U_k = 1 \text{ for all } k) = 0$. You can see this using a calculation using the fact that the U_k are independent

$$\Pr(U_k = 1, k = 1, \cdots, n) = p^n \ .$$

Clearly, if $U_k = 1$ for all k, then $U_k = 1$ for the first n "samples". This shows that

$$\Pr(U_k = 1 \text{ for all } k) \leq p^n \ , \quad \text{for all } n \ .$$

Zero is the only probability that is smaller than $\frac{1}{n}$ for all n.

Here's a more subtle example that's closer to the issue of Section 4. Suppose you estimate p from the sequence U_n. A natural estimator from the first n samples is

$$\begin{aligned} \widehat{p}_n &= \frac{1}{n} \# \{ U_k = 1 \mid 1 \leq k \leq n \} \\ \widehat{p}_n &= \frac{1}{n} \sum_{k=1}^{n} U_k \ . \end{aligned} \tag{11}$$

It is a theorem, see Subsection 3.2 that

$$\widehat{p}_n \to p \ , \quad \text{as } n \to \infty \ \text{ almost surely} \ . \tag{12}$$

This is an instance of the *strong law of large numbers*. Suppose A_p is the event that $\widehat{p}_n \to p$ as $n \to \infty$. Let B_p be the "complementary" event that the limit either does not exist or is not equal to p. Then *almost surely* means $\Pr_p(A_p) = 1$. The event B_p "never happens" in the $p-$world, which is $\Pr_p(B_p) = 0$.

The *probability space* of this example is the space of infinite sequences of zeros and ones. The $p-$measure is the probability distribution in which the U_k are independent and $U_k = 1$ with probability p. Let $q \neq p$ be a different probability between 0 and 1. We can define the $q-$measure in which the U_k are independent and $\Pr_q(U_k = 1) = q$. The strong law of large numbers for the $q-$measure says that $\widehat{p}_n \to q$ as $n \to \infty$, almost surely. This means that in the $q-$world, it is "almost sure" that $\lim_{n\to\infty} \widehat{p}_n \neq p$. Thus

$$\Pr_q(A_p) = 0 \ . \tag{13}$$

In this example we have two measures and a set A_p so that the $p-$measure of A_p is one (The *measure* of an event is its probability.) and the $q-$measure of A_p is zero.

3.2 Borel Cantelli

The *Borel Cantelli lemma* is a clever way to prove that limits happen almost surely. Here is a slightly non-standard way to explain the idea. It uses three ideas. First, if $S_n \geq 0$ is a sequence of non-negative numbers and if the infinite sum is finite, then the S_n have a limit

$$\sum_{n=1}^{\infty} S_n < \infty \implies \lim_{n\to\infty} S_n = 0 \; . \tag{14}$$

This statement is about any sequence of numbers. It is not probabilistic. Second, if $R \geq 0$ is a random variable that is allowed to take the value $R = \infty$, and if the expected value is finite, then R itself is finite almost surely

$$\mathrm{E}[R] < \infty \implies R < \infty \;\text{ almost surely} \; . \tag{15}$$

Third, you can exchange the order of summation and taking expected values for non-negative numbers

$$\sum_{n=1}^{\infty} \mathrm{E}[\, S_n] = \mathrm{E}\left[\sum_{n=1}^{\infty} S_n\right] \; . \tag{16}$$

There is a little more about these claims below, but first the consequence.

You can combine all three claims and get

$$S_n \geq 0 \; , \;\; \sum_{n=1}^{\infty} \mathrm{E}[\, S_n] < \infty \implies \lim_{n\to\infty} S_n = 0 \; , \;\text{ almost surely} \; . \tag{17}$$

Define R to be the sum of the non-negative numbers S_k:

$$R = \sum_{n=1}^{\infty} S_n \; .$$

The claim (16) tells you that if $\sum \mathrm{E}[\, S_n] < \infty$, then $\mathrm{E}[R] < \infty$. This is the right side of (16). Then (15) tells you that $R = \sum S_n < \infty$ almost surely. Finally, (14) tells you that $S_n \to 0$ as $n \to \infty$ almost surely. This (17) is a version of the Borel Cantelli lemma.

I hope the three claims (14), (15) and (16) are plausible. The first, (14), could be an exercise in an "$\epsilon - \delta$" introduction to analysis class. The second, (15) is a version of the inequality, for any $m > 0$, if $R \geq 0$, $\mathrm{E}[R] \geq m \Pr(R \geq m)$. This is a basic fact about probability measures (which are not defined in these notes). The third, (16) is a version of the *monotone convergence theorem.* The random variables

$$R_N = \sum_{n=1}^{N} S_n$$

are a monotone increasing sequence, which means that $R_{n+1} \geq R_n$. For any finite n,

$$\sum_{n=1}^{N} \mathrm{E}[\, S_n] = \mathrm{E}\left[\sum_{n=1}^{N} S_n\right] = \mathrm{E}[\, R_N] \; .$$

When $N \to \infty$, the left side converges to the left side of (16). The monotone convergence theorem says that the right side converges to the right side of (16).

We apply the Borel Cantelli lemma (17) and a calculation to prove the claim (12). The trick is to define

$$S_n = (\widehat{p}_n - p)^4 \ . \tag{18}$$

The power 4 on the right doesn't change the fact that $\widehat{p}_n - p \to 0$, but it does help the sum on the left of (17) converge. To see that, think about $a_n = \frac{1}{\sqrt{n}}$. The sum of a_n does not converge, but the sum of a_n^4 is the sum of $\frac{1}{n^2}$, which is finite.

$$a_n = \frac{1}{\sqrt{n}} \implies \sum_{n=1}^{\infty} a_n = \infty \ , \ \text{but} \ \sum_{n=1}^{\infty} a_n^4 = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \ .$$

The Borel Cantelli lemma is the conceptual framework of the proof. The technical part is the calculation which shows that

$$\mathrm{E}\Big[\ (\widehat{p}_n - p)^4 \Big] \le \frac{1}{n^2} \ . \tag{19}$$

The inequality (19) is the result of calculations that are "straightforward" (not subtle), but a little complicated. We start with

$$\widehat{p}_n - p = \frac{1}{n} \sum_{k=1}^{n} (U_k - p) \ .$$

The U_k on the right comes from the definition (11) of $\widehat{p}_n$. The p on the right comes from the p on the left and $\frac{1}{n} p = p$. Note that $\mathrm{E}[(U_k - p)] = 0$. We need a power of $\widehat{p}_n - p$. The trick for that is to express the power as a multiple sum. For example, the square is a double sum

$$\begin{aligned}
(\widehat{p}_n - p)^2 &= (\widehat{p}_n - p)\,(\widehat{p}_n - p) \\
&= \left(\frac{1}{n} \sum_{j=1}^{n} (U_j - p) \right) \left(\frac{1}{n} \sum_{k=1}^{n} (U_k - p) \right) \\
&= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} (U_j - p)\,(U_k - p) \ .
\end{aligned}$$

Therefore

$$\mathrm{E}\Big[\ (\widehat{p}_n - p)^2 \Big] = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathrm{E}[\,(U_j - p)\,(U_k - p)] \ . \tag{20}$$

If $j \ne k$, then $U_j - p$ is independent of $U_k - p$ and both have expected value zero. This gives

$$\mathrm{E}[\,(U_j - p)\,(U_k - p)] = 0 \ , \ \ \text{if } j \ne k \ .$$

The terms with $j = k$ have

$$\mathrm{E}\Big[\,(U_k - p)^2\Big] = \mathrm{var}(U_k - p) = p(1-p)\ .$$

There are n terms with $j = k$, so the sum is

$$\begin{aligned} \mathrm{E}\Big[\,(\widehat{p}_n - p)^2\Big] &= \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{var}(U_k - p) \\ &= \frac{1}{n^2}\, n\, p(1-p) \\ \mathrm{E}\Big[\,(\widehat{p}_n - p)^2\Big] &= \frac{p(1-p)}{n}\ . \end{aligned}$$

This calculation may be familiar from elementary probability because it is the calculation that shows that the variance of the sample mean is proportional to $\frac{1}{n}$. This rate of "decay to zero" is not fast enough to apply the Borel Cantelli lemma, because the sum of $\frac{1}{n}$ is infinite.

We get a better power of n using the fourth power instead of the square. The reasoning that led to (20) leads to

$$\mathrm{E}\Big[\,(\widehat{p}_n - p)^4\Big] = \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{E}[\,(U_i - p)\,(U_j - p)\,(U_k - p)\,(U_l - p)]\ .$$

As for the case with two $U - p$ terms, the expected value is zero if any of the indices i, j, k, l is not "matched" (equal to one of the others). For example, if $i \neq j$ and $i \neq k$ and $i \neq l$, then $U_i - p$ is independent of the other three terms and the expected value of the product is zero. You get terms with non-zero expectations in one of four ways. The first three ways involve distinct pairs and the fourth involves all four being equal

$$i = j\ ,\ \ k = l\ ,\ \ i \neq k\ ,\ \ n(n-1)\ \text{terms}\ ,\ \ \mathrm{E}\Big[\,(U_i - p)^2\,(U_k - p)^2\Big] = p^2(1-p)^2$$

$$i = k\ ,\ \ j = l\ ,\ \ i \neq j\ ,\ \ n(n-1)\ \text{terms}\ ,\ \ \mathrm{E}\Big[\,(U_i - p)^2\,(U_j - p)^2\Big] = p^2(1-p)^2$$

$$i = l\ ,\ \ j = k\ ,\ \ i \neq j\ ,\ \ n(n-1)\ \text{terms}\ ,\ \ \mathrm{E}\Big[\,(U_i - p)^2\,(U_j - p)^2\Big] = p^2(1-p)^2$$

$$i = j = k = l\ \ ,\ \ n\ \text{terms}\ ,\ \ \mathrm{E}\Big[\,(U_i - p)^4\Big] = p(1-p)(1-2p+2p^2)\ .$$

This leads to

$$\mathrm{E}\Big[\,(\widehat{p}_n - p)^4\Big] = \frac{3p^2(1-p)^2(n^2-n)}{n^4} + \frac{p(1-p)(1-2p+2p^2)n}{n^4}\ .$$

Most of the complexity of this expression is irrelevant. What matters here is whether the sum is finite. The first term on the right has n^2 in the numerator and n^4 in the denominator. The other term on the right has n "upstairs" and n^4 "downstairs". The sum of the first terms is, more or less, the sum of $\frac{1}{n^2}$ which does converge. This is a proof of the *strong law* of large numbers (12) for this case.

3.3 The Radon Nikodym theorem

The *Radon Nikodym theorem* is about when two probability measures are related by a likelihood ratio. A likelihood function $L(x)$ relates a $p-$ measure to a $q-$ measure if the expected value of any bounded continuous function $V(x)$ may be computed in either probability distribution:

$$\mathrm{E}_p[\,V(X)] = \mathrm{E}_q[\,V(X)L(X)]\ . \tag{21}$$

If this is true for "every" V, then we say the $p-$measure is *absolutely continuous* with respect to the $q-$measure. If $L(x) \neq 0$ for all x, then this relation may be turned around. If $W(x) = L^{-1}(x)V(x)$, then

$$\mathrm{E}_p\big[\,W(X)L^{-1}(W)\big] = \mathrm{E}_q[\,W(X)]\ . \tag{22}$$

If these are possible, then L is the *Radon Nikodym derivative* of the $p-$measure with respect to the $q-$measure.

The *Radon Nikodym theorem* says there is such an L unless some obvious necessary condition is violated. The necessary condition for (21) is, for any event A,

$$\mathrm{Pr}_q(A) = 0 \implies \mathrm{Pr}_p(A) = 0\ . \tag{23}$$

This rests on the following "obvious" fact that is a part of measure theory. If $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$, and if V is any other function, then

$$\mathrm{E}[\,V(X)\,\mathbf{1}(X)] = 0\ , \quad \text{if } \mathrm{Pr}(A) = 0\ . \tag{24}$$

If A is any event, and we take $V(x) = \mathbf{1}_A(x)$, then (21) says

$$\mathrm{Pr}_p(A) = \mathrm{E}[\mathbf{1}_A(X)L(X)]\ .$$

In this case $\mathrm{Pr}_q(A) = 0$ would imply that $\mathrm{Pr}_p(A) = 0$. The theorem (which is hard to prove) says that if the necessary condition (23) is satisfied, then there is a likelihood function L that makes (21) true. In that case, we say the $p-$measure is *absolutely continuous* with respect to the $q-$measure. If $q-$measure is also *absolutely continuous* with respect to the $p-$measure, then (22) is also true. In that case, we say the measures are *equivalent*. Equivalent measures are not the same. They can assign different probabilities to the same event and different expected values to the same function.

The opposite of absolutely continuous is *completely singular*. The measures are completely singular if there is an event A so that

$$\mathrm{Pr}_p(A) = 1\ , \quad \text{and} \quad \mathrm{Pr}_q(A) = 0\ . \tag{25}$$

The event $B = A^c$ goes the other way. We say that the $q-$measure puts "all its mass" in B while the $p-$measure puts none of its mass there. The sets A_p of Subsection 3.1 and (13) are an example.

4 Girsanov theory

Girsanov theory says which diffusion processes are equivalent to each other and which are not. When two diffusion processes are equivalent, it gives a formula for $L(X)$, which is *Girsanov's formula*. The answer is, skipping the fine print, that two diffusions are equivalent if they have the same infinitesimal variance and they are completely singular with respect to each other otherwise. You can change the infinitesimal mean (the drift) but not the infinitesimal variance (quadratic variation) with a likelihood function change of measure.

You can see these facts explicitly for Brownian motion. We define $dX_t = X_{t+dt} - X_t$. Suppose there is a $W-$world where X_t is just Brownian motion, with zero infinitesimal mean and unit infinitesimal variance:

$$\mathrm{E}_W\left[\, dX_t \mid X_{[0,t]}\right] = 0$$
$$\mathrm{E}_W\left[\, (dX_t)^2 \mid X_{[0,t]}\right] = dt \; .$$

In the $Z-$world, there is an adapted infinitesimal mean function a_t so that

$$\mathrm{E}_Z\left[\, dX_t \mid X_{[0,t]}\right] = a_t dt$$
$$\mathrm{E}_Z\left[\, (dX_t)^2 \mid X_{[0,t]}\right] = dt \; .$$

The likelihood function that changes measure from W to Z is given by Girsanov's formula

$$L(X) = e^{\int_0^T a_t dX_t} e^{-\frac{1}{2}\int_0^T a_t^2 dt} \; . \tag{26}$$

If $V(x)$ is any path functional defined on $[0,T]$, then

$$\mathrm{E}_Z\left[\, V(X_{[0,T]})\right] = \mathrm{E}_W\left[\, V(X_{[0,T]})L(X_{0,T]})\right] \; . \tag{27}$$

5 Exercises

1. Calculate the $q-$world variance of $e^{\frac{1}{2}\mu}e^{\mu X}V(X)$ as a function of μ and a, in (10). What value of μ given the smallest variance, for $a > 0$ and large? What is the ratio of the $p-$world standard deviation to the optimal $q-$world standard deviation?

2. Consider the example functional that has backward equation (7). Suppose X_t is Brownian motion and consider the random variable

$$Z_{x,t} = \int_t^T X_s \, ds \; , \quad \text{conditioned on } X_t = x \; .$$

Show that $Z_{x,t}$ is Gaussian and identify the mean and variance. Use that information to get $f(x,t) = \mathrm{E}\left[\, e^{Z_{x,t}}\right]$. This should agree with the answer found by the ansatz "method".

3. If X_t is Brownian motion but $V(x) = x^2$, then the functional is

$$\int_t^T X_t^2 \, dt \; \neq \; \text{Gaussian} \; .$$

   Look for a solution using the ansatz method that involves $e^{A(t)x^2}$.

4. Consider the Brownian motion with positive constant drift and positive starting point

$$dX_t = adt + dW_t \; , \;\; X_0 = x_0 \; , \;\; a > 0 \; , \;\; x_0 > 0 \; .$$

   Let τ be the first hitting time when $X_t = 0$. Be aware that $\Pr(\tau < \infty) < 1$, which means it is possible (there is a non-zero probability) that $X_t > 0$ for all $t \geq 0$. Make a histogram of τ for $a = .5$ and $x_0 = .3$. You need to specify a final time T. Do this by trial and error, so that there are not many hits after time T. In the code, these must not be "hard wired", which means, for example, that the code should use a variable `a` that is assigned `a = .5`. Reproduce the histogram by simulating Brownian motion directly and putting in the Girsanov change of measure weight. For this, you have to work with a *weighted histogram*, where values of τ come with weights. Part of the exercise is to figure out how to do this. When you're computing the Girsanov factor, you need only integrate up to τ, not the final time T. Why?