Chapter 10

Feynam-Kac formulae

In this Chapter we will consider relationships between diffusion processes and certain second order partial differential equations (PDE).

In the case of diffusion processes, the corresponding infinite dimensional equations are partial differential equations (PDE). But we will not restrict our attention to diffusion processes. For our purposes, we can assume that $X^{(t)}$ is either continuous or is right continuous with left hand limits (i.e. the limit $X^{(t^-)} = \lim_{s \uparrow t} X^{(s)}$ exists), is adapted to some filtration $\mathcal{F}_t$, and is the solution to the martingale problem

$$f(X^{(t)}) = f(x) + \int_0^t \mathcal{L}^{(s)} f(X^{(s)}) ds + M_f^{(t)} \tag{10.1}$$

where $M_f^{(t)}$ depends on f and is an $\mathcal{F}_t$-martingale. We can assume that $M_f^{(t)}$ is either continuous (in which case if it is non-zero it must have infinite first variation), or it has finite first variation. For any particular generator $\mathcal{L}$, the class of test functions f will have to be restricted (e.g. $\mathcal{C}^2$ functions when $\mathcal{L}$ is a diffusion operator). We will assume that any function to which we apply $\mathcal{L}$ satisfies the required restrictions. The process $X^{(t)}$ is a semi-martingale, that is, we can write

$$X^{(t)} = A^{(t)} + N^{(t)}$$

where $A^{(t)}$ is a (possibly discontinuous) finite variation process and $N^{(t)}$ is a martingale. The family of such $X^{(t)}$ includes both diffusions and jump Markov processes, as well as combinations of the two.

Note that for such a process we can define the quadratic variation and stochastic integral with respect to X exactly as we did before. The Itô integration by parts formula for two semi-martingales, X and Y, becomes

$$X^{(t)}Y^{(t)} - X^{(0)}Y^{(0)} = \int_0^t X^{(s^-)}dY^{(s)} + \int_0^t Y^{(s^-)}dX^{(s)} + \sum_{S_j \le t} \triangle_{S_j^-}^{S_j}(XY)$$
$$+ [X,Y]^{(t)} - \sum_{S_j \le t} X^{(S_j^-)}\triangle_{S_j^-}^{S_j} Y - \sum_{S_j \le t} Y^{(S_j^-)}\triangle_{S_j^-}^{S_j} X - \sum_{S_j \le t} \triangle_{S_j^-}^{S_j} X \triangle_{S_j^-}^{S_j} Y$$

where the discontinuities of the process $X^{(t)}Y^{(t)}$ occur at times S_j. From this formula we can see that if Y happens to be a continuous process with finite variation, the Itô product rule again reduces to the classical product rule, a fact that will be used repeatedly below.

10.1 The Fokker-Planck and Kolmogorov equations

We will begin by considering the parabolic, backwards in time equation

$$-\left(\partial_r + \mathcal{L}^{(r)} - V^{(r)}\right) u = 0, \; r \in [0,t], \qquad u(t,x) = f(x) \tag{10.2}$$

where the term $V^{(r)}u$ represents pointwise multiplication, i.e. $V^{(r)}u(r,x) = V(r,x)u(r,x)$. In the case that $\mathcal{L}^{(t)}$ is the generator of a diffusion process, i.e.

$$\mathcal{L}^{(t)} f(x) = (\nabla f(x))\, b(t,x) + \frac{1}{2}\text{trace}\left(a(t,x)D^2 f(x)\right) \tag{10.3}$$

the equation (10.2) is a second order parabolic terminal value PDE problem. The function V in (10.2) is called the potential. In (10.3) it is assumed that

$$a(t,x) = (\sigma(t,x))\,(\sigma(t,x))^{\mathrm{T}}.$$

We will assume that the drift and diffusion coefficients b, σ, and the potential term V are chosen so that this PDE equation has a unique and twice continuously differentiable (i.e. classical) solution u.

Equations of the form in (10.2) appear in many contexts. This is particularly apparent when we note that, if $\mathcal{L}$ and V, are independent of time, the solution to (10.2), u, satisfies $u(r,x) = v(t-r,x)$ for $r \in [0,t]$, where v solves the equation

$$(\partial_r - \mathcal{L} + V)\, v = 0,\ r \geq 0, \qquad v(0,x) = f(x). \tag{10.4}$$

For example, when $\mathcal{L}$ is of the form in (10.3) with $b = V = 0$, and σ constant, (10.4) describes the temperature of an object initially at temperature $f(x)$ at position x. Also when $\mathcal{L}$ is of the form in (10.3), when $V \neq 0$, $b = 0$, and $\sigma = 1$, (10.2) is called the imaginary time Schrödinger equation, the solutions of which, we will see later, can be used to find the ground state energy of a molecular system.

We will find a representation of the solution to u to (10.2) in terms of the solution $X^{(t)}$ to the martingale problem (10.1). Notice that if

$$Y^{(r)} = u(r, X^{(r)}) e^{-\int_{t_0}^r V(s,X^{(s)})ds}$$

then

$$dY^{(r)} = e^{-\int_{t_0}^r V(s,X^{(s)})ds} \left(\partial_r + \mathcal{L}^{(r)} - V^{(r)}\right) u(r, X^{(r)}) dr \\ + e^{-\int_{t_0}^r V(s,X^{(s)})ds} dM_u^{(r)}$$

which, since the first term vanishes when u solves (10.2), shows that $Y^{(r)}$ is a martingale on $[t_0, t]$. Therefore, we must have that

$$u(t_0, x) = \mathbf{E}\left[Y^{(t_0)}\right] = \mathbf{E}\left[Y^{(t)}\right] = E_{t_0,x}\left[f(X^{(t)}) e^{-\int_{t_0}^t V(s,X^{(s)})ds}\right] \tag{10.5}$$

where, as in Chapter ?? $E_{t,x}$ denotes the expectation with respect to the distribution of the process $X^{(r)}$ for $r \geq t$ assuming that $X^{(t)} = x$. As a consequence of formula (10.5), when $\mathcal{L}$ and V are independent of time we find the representation

$$v(t,x) = \mathbf{E}_x\left[f(X^{(t)}) e^{-\int_0^t V(X^{(s)})ds}\right] \tag{10.6}$$

for v solving (10.4).

Now we will consider representation of equations involving an adjoint of the Kolmogorov operator, the Fokker–Planck operator $\left(\mathcal{L}^{(t)}\right)^*$ defined by the relation

$$\int f(x) \left(\mathcal{L}^{(t)}\right)^* g(x) dx = \int g(x) \mathcal{L}^{(t)} f(x) dx$$

for any appropriate test functions f and g. When $\mathcal{L}$ is a diffusion operator as in (10.3),

$$\left(\mathcal{L}^{(t)}\right)^* \rho = -\mathrm{div}(\rho\, b) + \frac{1}{2}\sum_{i,j} \partial_{ij}^2(\rho\, a_{ij}).$$

Suppose that $\rho(t,x)$ solves the equation

$$\left(\partial_t - \left(\mathcal{L}^{(t)}\right)^* + V^{(t)}\right)\rho = 0,\ t \geq 0, \qquad \rho(0,x) = \mu(x) \tag{10.7}$$

for some probability density μ. Then, for u solving (10.2) consider the function

$$\varphi^{(r)} = \int u(r,x)\rho(r,x)dx,\ r \in [t_0, t]$$

Clearly, $\varphi^{(0)} = \int u(t_0,x)\mu(x)dx$ and $\varphi^{(t)} = \int f(x)\rho(t,x)dx$, and upon taking the derivative of φ we find that

$$\begin{aligned}\frac{d}{dr}\varphi^{(r)} = &-\int \rho(r,x)\left(\mathcal{L}^{(r)} - V^{(r)}\right)u(r,x)dx \\ &+\int u(r,x)\left(\left(\mathcal{L}^{(r)}\right)^* - V^{(r)}\right)\rho(r,x)dx.\end{aligned}$$

By the definition of the adjoint (and since the operator $u \to V^{(r)}u$ is self-adjoint), the two terms on the right hand side of the last display cancel each other so that $\varphi^{(0)} = \varphi^{(t)}$, i.e.

$$\int f(x)\rho(t,x)dx = \int u(0,x)\mu(x)dx = E_\mu\left[f(X^{(t)})e^{-\int_0^t V(s,X^{(s)})ds}\right]. \tag{10.8}$$

where the subscript μ on the expectation indicates that $X^{(0)}$ is drawn from μ. In other words the solution to (10.7) will remain non-negative and the normalized solution

$$\eta(t,x) = \frac{\rho(t,x)}{\int \rho(t,x)dx} \tag{10.9}$$

is the density of the random variable $X^{(t)}$ under the probability Q whose change of measure with respect to the original distribution P restricted to $[0,t]$ is

$$\left.\frac{dQ}{dP}\right|_t = \frac{e^{-\int_0^t V(s,X^{(s)})ds}}{E_\mu\left[e^{-\int_0^t V(X^{(s)})ds}\right]}$$

10.2 Boundary value problems

We now move on to finding a stochastic representation for the boundary value problem

$$-\left(\mathcal{L}-V\right)u = g \text{ for } x \in D, \qquad u(x) = f(x) \text{ for } x \notin D \tag{10.10}$$

where D is a bounded open set with smooth boundary. Here we assume that $\mathcal{L}$ and V are independent of time. Set the escape time of $X^{(t)}$ from D to be

$$\tau_{D^c} = \inf\left\{t \geq 0 : \, X^{(t)} \notin D\right\}.$$

Note that right continuity of $X^{(t)}$ implies that $X^{(\tau_{D^c})} \notin D$. We will assume that $\tau_{D^c} < \infty$ with probability 1. Define

$$Y^{(t)} = u(X^{(t)})e^{-\int_0^t V(X^{(s)})ds} + \int_0^t g(X^{(s)})e^{-\int_0^s V(X^{(r)})dr}ds.$$

We find that

$$dY^{(t)} = e^{-\int_0^t V(X^{(s)})ds} dM_u^{(t)},$$

in other words, that $Y^{(t)}$ is again a martingale. Hence we find that

$$u(x) = E_x\left[f(X^{(\tau_{D^c})})e^{-\int_0^{\tau_{D^c}} V(X^{(s)})ds} + \int_0^{\tau_{D^c}} g(X^{(s)})e^{-\int_0^s V(X^{(r)})dr}ds\right].$$

Now imagine that $f(x) = \mathbf{1}_B(x)$ where $D = (A \cup B)^c$ for disjoint sets A and B. Suppose further that $V = g = 0$. We have just shown that, in this case, the solution to (10.10), u, satisfies

$$u(x) = P_x\left[\tau_A > \tau_B\right]$$

where τ_A and τ_B are the first hitting times of A and B. The function $P_x\left[\tau_A > \tau_B\right]$ is called the forward committor (or just committor if $X^{(t)}$ is reversible) function and is an important object in non-equilibrium statistical mechanics, used to describe properties of the transitions of $X^{(t)}$ out of A and into B. Chemical reactions and molecular conformational changes are important examples of just that kind of transition.

Setting $f = V = 0$ and $g = 1$ we find that the solution to (10.10) is the mean escape time from D, $u(x) = E_x\left[\tau_{D^c}\right]$. In fact, by taking $V(x) = \lambda$, $f = 1$, and $g = 0$ we find that the solution to (10.10) is

$$u(x;\lambda) = E_x\left[e^{\lambda\tau_{D^c}}\right],$$

i.e. $u(x;\lambda)$ is the value of the Laplace transform of the distribution of the escape time at parameter λ. If we know $u(x;\lambda)$ of all values of λ, we can compute any moment of τ_{D^c}.

10.3 Dominant eigenproblems

In several applications, the goal is to find the smallest eigenvalue, λ_0, of the operator

$$\mathcal{H} = -\mathcal{L}^* + V$$

and its corresponding eigenfunction ψ_0. We will assume that the eigenvector corresponding to λ_0 is unique. Because λ_0 is the smallest eigenvector, we expect that if $\rho(t,x)$ solves the Fokker Planck equation (10.7) then the normalized solution η in (10.9) should converge to ψ. For example, if the eigenvectors of $\mathcal{H}$ form a countable basis for the domain of $\mathcal{L}^*$ then we can imagine expand the initial density μ as

$$\mu(x) = \sum_{j=0}^{\infty} \alpha_j \psi_j(x)$$

for some sequence α_j and where the eigenpairs of $\mathcal{H}$ are (λ_j, ψ_j). In this case the solution to (10.7) becomes

$$\rho(t,x) = \sum_{j=0}^{\infty} \alpha_j e^{-\lambda_j t} \psi_j(x)$$

so that $\eta(t,x)$ does indeed converge to $\psi_0(x)$ (normalized to have integral equal to 1). When $\eta(t,x)$ converges to $\psi_0(x)$ for large t, we obtain the stochastic representation

$$\int f(x)\psi_0(x)dx = \lim_{t\to\infty} \frac{E_x\left[f(X^{(t)})e^{-\int_0^t V(X^{(s)})ds}\right]}{E_x\left[e^{-\int_0^t V(X^{(s)})ds}\right]}.$$

Now consider the equation solved by η,

$$\partial_t \eta(t,x) = -\mathcal{L}^* \eta(t,x) + (V(x) - \lambda_\eta(t))\eta(t,x), \qquad \eta(0,x) = \mu(x)$$

where

$$\lambda_\eta(t) = \int V(x)\eta(t,x)dx.$$

The stochastic representation for this equation yields

$$1 = \int \eta(t,x)dx = E_x\left[e^{-\int_0^t (V(X^{(s)}) - \lambda_\eta(s))ds}\right],$$

from which we find that

$$\int_0^t \lambda_\eta(s)ds = -\log E_x\left[e^{-\int_0^t V(X^{(s)})ds}\right].$$

If $\eta(t,x)$ converges to $\psi_0(x)$ as t increases, then we expect that

$$\lim_{t\to\infty} \lambda_\eta(t) = \int V(x)\psi_0(x)dx.$$

On the other hand, from the eigenequation for (λ_0, ψ_0) we find that

$$\begin{aligned}
\lambda_0 &= \int \mathcal{H}\psi_0(x)dx \\
&= -\int \mathcal{L}^*\psi_0(x)dx + \int V(x)\psi_0(x)dx \\
&= -\int \psi_0(x)\mathcal{L}1(x)dx + \int V(x)\psi_0(x)dx \\
&= \int V(x)\psi_0(x)dx
\end{aligned}$$

so that

$$\lim_{t\to\infty} \lambda_\eta(t) = \lambda_0.$$

From these expressions we can conclude that

$$\lambda_0 = \lim_{t\to\infty} \frac{1}{t}\int_0^t \lambda_\eta(s)ds = \lim_{t\to\infty} -\frac{1}{t}\log E_x\left[e^{-\int_0^t V(X^{(s)})ds}\right].$$

10.4 A stochastic control representation for diffusions

In this section we assume that $X^{(t)}$ is the solution to the SDE

$$dX^{(t)} = b(t, X^{(t)})\, dt + \sigma(t, X^{(t)})\, dW, \qquad X^{(0)} = x$$

and has generator $\mathcal{L}$. From what we have already learned, the function $u(r, x) = E_{r,x}\left[e^{-g(X^{(t)}) - \int_r^t V(X^{(s)})ds}\right]$ solves the backward PDE

$$(\partial_r + \mathcal{L} - V)\, u = 0, \qquad u(t, x) = e^{-g(x)}.$$

Consider then the PDE solved by the function $w(r, x) = -\log u(r, x)$. For a diffusion, one can easily verify that

$$e^w \mathcal{L}[e^{-w}] = -\mathcal{L}w + \frac{1}{2}\|\sigma^{\mathrm{T}}\nabla^{\mathrm{T}} w\|_2^2$$

so that w solves the second order Hamilton–Jacobi equation

$$-\partial_r w - \mathcal{H}[w] + V = 0, \qquad w(t, x) = g(x). \tag{10.11}$$

where we have introduced the non-linear operator

$$\mathcal{H}[w] = \mathcal{L}w - \frac{1}{2}\|\sigma^{\mathrm{T}}\nabla^{\mathrm{T}} w\|_2^2.$$

The solution to this equation has the stochastic control representation

$$w(r, x) = \inf_{U \in \mathcal{A}[r,t]} E_{r,x}\left[\int_r^t \left(\frac{1}{2}\|U^{(s)}\|_2^2 + V(X_U^{(s)})\right) ds + g(X_U^{(t)})\right] \tag{10.12}$$

where the process $X_U^{(t)}$ solves the SDE

$$dX_U^{(t)} = b(t, X_U^{(t)})\, dt + \sigma(t, X_U^{(t)}) U^{(t)}\, dt + \sigma(t, X_U^{(t)})\, dW, \qquad X_U^{(0)} = x$$

and the infimum can be taken over all absolutely continuous processes on $[r, t]$ adapted to the filtration generated by W. To see that this is the case first recall that From Girsanov's formula, we know that we can write

$$w(r, x) = -\log E_{r,x}\left[e^{-g(X_U^{(t)}) - \int_r^t V(X_U^{(s)})ds} e^{-\int_r^t U^{(s)} dW^{(s)} - \frac{1}{2}\int_r^t \|U^{(s)}\|_2^2 ds}\right]$$

and, applying Jensen's inequality, we find that, for any U,

$$w(r,x) \le E_{r,x}\left[\int_r^t \left(\frac{1}{2}\|U^{(s)}\|_2^2 + V(X_U^{(s)})\right) ds + g(X_U^{(t)})\right]$$

Now consider U of the form

$$\hat{U}^{(t)} = v(t, X_{\hat{U}}^{(t)}).$$

For such a U, the action of the generator of $X_U^{(t)}$ becomes

$$\mathcal{L}_U f = \mathcal{L} f + \nabla_x f \,\sigma v.$$

Now fix $t_0 < t$, set $X_U^{(t_0)} = x$, and define

$$Y^{(r)} = u(r, X_U^{(r)}) e^{-\int_{t_0}^r V(X_U^{(s)})ds - \int_{t_0}^r v(s, X_U^{(s)}) dW^{(s)} - \frac{1}{2}\int_{t_0}^r \|v(s,X_U^{(s)})\|_2^2 ds}$$

we find that

$$\begin{aligned} dY &= Y\Big((\partial_r + \mathcal{L}_U) u\, dr + \nabla_x u\, \sigma\, dW - u\,(V\, dr + v\, dW) - \nabla_x u\, \sigma v dr\Big) \\ &= Y(\nabla_x u\, \sigma - uv)\, dW. \end{aligned}$$

For the particular choice,

$$v = \frac{\sigma^{\mathrm{T}} \nabla_x^{\mathrm{T}} u}{u} = -\sigma^{\mathrm{T}} \nabla^{\mathrm{T}} w,$$

we see arrive at the equality

$$e^{-g(X_U^{(t)}) - \int_{t_0}^t V(X_U^{(s)})ds - \int_{t_0}^t v(s, X_U^{(s)}) dW^{(s)} - \frac{1}{2}\int_{t_0}^t \|v(s,X_U^{(s)})\|_2^2 ds} = u(t_0, x)$$

so that, for this choice of U,

$$w(t_0,x) = E_{t_0,x}\left[\int_{t_0}^t \left(\frac{1}{2}\|U^{(s)}\|_2^2 + V(X_U^{(s)})\right) ds + g(X_U^{(t)})\right]$$

and the stochastic control representation in (10.12) is confirmed. In fact, (10.12) holds for any expectation of the form $E\left[e^{-F(X)}\right]$ where F is any bounded and measurable functional of the path of X, though the proof of that result is more complicated.

Now notice that, for $t_0 < r < t$,

$$\begin{aligned} u(t_0, x) &= E_{t_0,x}\left[e^{-g(X^{(t)})-\int_{t_0}^t V(X^{(s)})ds}\right] \\ &= E_{t_0,x}\left[E_{r,X^{(r)}}\left[e^{-g(X^{(t)})-\int_r^t V(X^{(s)})ds}\right] e^{-\int_{t_0}^r V(X^{(s)})ds}\right] \\ &= E_{t_0,x}\left[u(r, X^{(r)})e^{-\int_r^t V(X^{(s)})ds}\right] \end{aligned}$$

which we can also write as

$$w(t_0, x) = -\log E_{t_0,x}\left[e^{-w(r,X^{(r)})-\int_r^t V(X^{(s)})ds}\right].$$

Our argument above then implies that

$$w(t_0, x) = \inf_{U\in\mathcal{A}[t_0,r]} E_{t_0,x}\left[\int_{t_0}^r \left(\frac{1}{2}\|U^{(s)}\|_2^2 + V(X_U^{(s)})\right) ds + w(r, X_U^{(r)})\right]. \tag{10.13}$$

This last expression is referred to as a dynamic programming principle and can be used to directly (without further reference to u) verify the Hamilton–Jacobi equation (10.11) satisfied by w.

10.5 The Doob h-transform and importance sampling

We now shift our focus to estimating some of the expectations that we have come across in this chapter. The practical utility of the Feynman–Kac formulae that we have derived is that they give a representation of the solutions to important infinite dimensional equations (e.g. backward Kolmogorov and Fokker–Planck equations) in terms of an expectation that can be estimated by Monte Carlo. This is useful because it is generally only possible to directly approximate the solution to these infinite dimensional equations when their independent variables are very low dimensional (rarely higher than 4 dimensions). In contrast, as long as the variance of the random variable $f(X^{(t)})e^{-\int_0^t V(s,X^{(s)})ds}$ does not increase with dimension, the cost to achieve a fixed accuracy with the estimator

$$\overline{f}_{V,N}^{(t)} = \frac{1}{N}\sum_{j=0}^{N-1} f(X^{(t,j)})e^{-\int_0^t V(s,X^{(s,j)})ds}$$

with N independent copies $X^{(t,j)}$ of $X^{(t)}$, depends only weakly on the dimension of $X^{(t)}$ (through the cost of integrating $X^{(t)}$).

Unfortunately, the variance of this estimator is often extremely high. For example, if $X^{(t)}$ is a Brownian motion, $f(x) = 1$, and $V(s,x) = x$, then, since $\int_0^t W^{(s)} ds$ is an $N(0, t^3/3)$ random variable,

$$\mathbf{var}\left(\overline{1}^{(t)}_{x,N}\right) = \frac{e^{\frac{t^3}{3}}}{N}$$

which grows very quickly with t.

The simplest way to try to reduce the variance of this estimator is to use the importance sampling estimator

$$\widetilde{f}^{(t)}_{V,N} = \frac{1}{N}\sum_{j=0}^{N-1} f(Y^{(t,j)}) e^{-\int_0^t V(s,Y^{(s,j)})ds} Z^{(t,j)}$$

where $Y^{(s,j)}$ are independent realizations of some alternative process, $Y^{(t)}$, and $Z^{(t,j)}$ are the corresponding realizations of the change of measure so that $\widetilde{f}^{(t)}_{V,N}$ remains an unbiased estimator. Our discussion of the optimal importance sampling estimator from Chapter ?? suggests that we replace the original path distribution P by the path distribution with change of measure (on $[0,t]$),

$$\left.\frac{dQ}{dP}\right|_t = \frac{|f(X^{(t)})| e^{-\int_0^t V(s,X^{(s)})ds}}{\mathbf{E}_x\left[|f(X^{(t)})| e^{-\int_0^t V(s,X^{(s)})ds}\right]}. \tag{10.14}$$

When f is non-negative, for example, the estimator $\widetilde{f}^{(t)}_{V,N}$ with independent samples $Y^{(t,j)}$ drawn from Q and

$$Z^{(t,j)} = \frac{\mathbf{E}_x\left[f(X^{(t)}) e^{-\int_0^t V(s,X^{(s)})ds}\right]}{f(Y^{(t,j)}) e^{-\int_0^t V(s,Y^{(s,j)})ds}}$$

has zero variance.

Below we will see that the measure Q is associated with a transition density whose evolution we can characterize exactly. To see that, first suppose that

$\rho(r, y \,|\, s, x)$ solves the forward equation (10.7) for $r \geq s$ with initial condition $\rho(s, y \,|\, s, x) = \delta(y - x)$. In other words, suppose that

$$\int f(y) p(r, y \,|\, s, x) dy = E_{s,x}\left[f(X^{(r)}) e^{-\int_s^r V(z, X^{(z)}) dz} \right]$$

for any test function f. For the particular test function f of interest, we define a new density by setting

$$q(r, y \,|\, s, x) = \frac{u(r, y) p(r, y \,|\, s, x)}{\int u(r, y) p(r, y \,|\, s, x) dy}$$

where u solves the backward equation (10.2), i.e.

$$u(r, x) = E_{r,x}\left[|f(X^{(t)})| e^{-\int_r^t V(s, X^{(s)}) ds} \right].$$

As one can easily check, q does satisfy the Chapman-Kolmogorov equation

$$q(r, y \,|\, s, x) = \int q(r, y \,|\, s', z) q(s', z \,|\, s, x)\, dz$$

for any $s' \in [s, r]$ as required for a valid Markov transition density. This would not be the case, for example, if we were to replace $u(r, y)$ in the equation for q by $|f(y)|$. This transformation of p is a slight generalization of what is usually referred to as Doob's h-transformation. Notice that

$$q(t, y \,|\, 0, x) = \frac{|f(y)| p(t, y \,|\, 0, x)}{\int |f(y)| p(t, y \,|\, 0, x) dy} = \frac{|f(y)| p(t, y \,|\, 0, x)}{E_x\left[|f(X^{(t)})| e^{-\int_0^t V(s, X^{(s)}) ds} \right]} \tag{10.15}$$

which is indeed the density corresponding to Q in (10.14).

We turn now to characterizing the evolution of the density q. Taking its derivative with respect to time we find that

$$\begin{aligned} \partial_r q &= \frac{p \partial_r u + u \partial_r p}{\int up\, dy} - q \frac{\int \partial_r (up)\, dy}{\int up\, dy} \\ &= \frac{p(-\mathcal{L}u + Vu) + u(\mathcal{L}^* p - Vp)}{\int up\, dy} - q \frac{\int \partial_r (up)\, dy}{\int up\, dy}. \end{aligned}$$

Cancelling terms we find that

$$\partial_r q = -\frac{q}{u} \mathcal{L} u + u \mathcal{L}^* \left(\frac{q}{u} \right) + q \int \left(\frac{q}{u} \mathcal{L} u - u \mathcal{L}^* \left(\frac{q}{u} \right) \right) dy.$$

Inspecting the integral term we find that it vanishes because

$$\int u\mathcal{L}^*\left(\frac{q}{u}\right)\,dy = \int \frac{q}{u}\mathcal{L}u\,dy$$

so that, finally, we end up with the equation

$$\partial_r q = \tilde{\mathcal{L}}^* q, \quad r \geq s, \qquad q(s, y \,|\, s, x) = \delta(y - x) \tag{10.16}$$

where we have introduced the new operator

$$\tilde{\mathcal{L}}^* q = -\frac{q}{u}\mathcal{L}u + u\mathcal{L}^*\left(\frac{q}{u}\right).$$

The adjoint of this operator is

$$\tilde{\mathcal{L}} f = -\frac{f}{u}\mathcal{L}u + \frac{1}{u}\mathcal{L}(f\,u).$$

The facts that this operator is linear and that $\tilde{\mathcal{L}}1 = 0$ relates to our interpretation of q as the transition operator for a Markov process.

The exact interpretation of (10.16) depends on the particular form of the generator. For example, it is not obvious that if $\mathcal{L}$ corresponds to a process of a certain type (e.g. a diffusion) the transition density q will still correspond to a process of the same type. In the particular case of diffusions this is, however, true. To see this notice that if $\mathcal{L}$ generates a diffusion,

$$\begin{aligned}
\mathcal{L}^*\left(\frac{q}{u}\right) &= -\sum_j \partial_j\left(\frac{q\,b}{u}\right) + \frac{1}{2}\sum_{i,j}\partial_{ij}^2\left(\frac{q\,a_{ij}}{u}\right) \\
&= -\frac{1}{u}\sum_j\left(\partial_j(q\,b) - q\,b\,\frac{\partial_j u}{u}\right) + \frac{1}{2\,u}\sum_{i,j}\partial_{ij}^2(q\,a_{ij}) \\
&\qquad - 2\partial_j(q\,a_{ij})\frac{\partial_i u}{u} - q\,a_{ij}\frac{\partial_{ij}^2 u}{u} + 2q\,a_{ij}\frac{\partial_i u\,\partial_j u}{u^2} \\
&= \frac{1}{u}\mathcal{L}^* q + \frac{q}{u^2}\mathcal{L}u - \frac{1}{u}\sum_{i,j}\partial_j(q\,a_{ij})\frac{\partial_i u}{u} + q\,a_{ij}\frac{\partial_{ij}^2 u}{u} - q\,a_{ij}\frac{\partial_i u\,\partial_j u}{u^2} \\
&= \frac{1}{u}\mathcal{L}^* q + \frac{q}{u^2}\mathcal{L}u - \frac{1}{u}\sum_{i,j}\partial_j(q\,a_{ij})\frac{\partial_i u}{u} + q\,a_{ij}\partial_{ij}^2(\log u) \\
&= \frac{1}{u}\mathcal{L}^* q + \frac{q}{u^2}\mathcal{L}u - \frac{1}{u}\mathrm{div}\left(q\,a\,\nabla_x^{\mathrm{T}}\log u\right)
\end{aligned}$$

so that

$$\partial_r q = \mathcal{L}^* q - \operatorname{div}\left(q\, a\, \nabla_x^{\mathrm{T}} \log u\right).$$

But the right hand side of the last display is exactly the adjoint of the generator corresponding to the SDE

$$dY^{(t)} = b(t, Y^{(t)})\, dt + \sigma(t, Y^{(t)})\, v(t, Y^{(t)})\, dt + \sigma(t, Y^{(t)})\, dW^{(t)}, \qquad Y^{(0)} = x$$

with

$$v = \sigma^{\mathrm{T}}\, \nabla_x^{\mathrm{T}} \log u.$$

In retrospect, this is not a surprise at all. In the last section we arrived at exactly the same function v when deriving a control process $U^{(s)} = v(s, X_U^{(s)})$ that made

$$e^{-g(X_U^{(t)}) - \int_{t_0}^t V(X_U^{(s)}) ds - \int_{t_0}^t v(s, X_U^{(s)}) dW^{(s)} - \frac{1}{2}\int_{t_0}^t \|v(s, X_U^{(s)})\|_2^2 ds}$$

always equal to $u(t_0, x)$.

Now suppose that in the definition of q in (10.15) we replace $u(r, x)$ by φ_0, the eigenvector of $-\mathcal{L} + V$ corresponding to its smallest eigenvalue, i.e.

$$-\left(\mathcal{L} - V\right) \varphi_0 = \lambda_0 \varphi_0.$$

Following the steps above, we find that the equation solved by q is now

$$\begin{aligned}
\partial_r q &= \varphi_0 \left(\mathcal{L}^* - V\right) \left(\frac{q}{\varphi_0}\right) - q \int \varphi_0 \left(\mathcal{L}^* - V\right) \left(\frac{q}{\varphi_0}\right) dy \\
&= \varphi_0 \left(\mathcal{L}^* - V\right) \left(\frac{q}{\varphi_0}\right) - q \int (\mathcal{L} - V)\varphi_0 \left(\frac{q}{\varphi_0}\right) dy \\
&= \varphi_0 \left(\mathcal{L}^* - V\right) \left(\frac{q}{\varphi_0}\right) + \lambda_0 q
\end{aligned}$$

where in the last equality we have remembered that $\int q\, dy = 1$. This is again a linear evolution of the form

$$\partial_r q = \tilde{\mathcal{L}}^* q$$

where now

$$\begin{aligned}
\tilde{\mathcal{L}} f &= \frac{1}{\varphi_0}\, \mathcal{L}(f\, \varphi_0) - (V - \lambda_0) f \\
&= \frac{1}{\varphi_0}\, \mathcal{L}(f\, \varphi_0) - \frac{f}{\varphi_0} \mathcal{L}\varphi_0
\end{aligned}$$

which again satisfies $\mathcal{L}1 = 0$. For example, when $\mathcal{L}$ is a diffusion operator, our considerations above imply that $\tilde{\mathcal{L}}$ is the generator of the diffusion

$$dY^{(t)} = b(t, Y^{(t)})\, dt + \sigma(t, Y^{(t)})\, v(t, Y^{(t)})\, dt + \sigma(t, Y^{(t)})\, dW^{(t)}, \qquad Y^{(0)} = x$$

with

$$v = \sigma^{\mathrm{T}}\, \nabla_x^{\mathrm{T}} \log \varphi_0.$$

Note that if $\tilde{\psi}_0$ is the invariant probability density corresponding to $\tilde{\mathcal{L}}$ then

$$-\left(\mathcal{L}^* - V\right) \frac{\tilde{\psi}_0}{\varphi_0} = \lambda_0 \frac{\tilde{\psi}_0}{\varphi_0},$$

i.e.

$$\psi_0 = \frac{\tilde{\psi}_0}{\varphi_0}.$$

The key to utility of these observations in random approaches to approximate the eigenvector, ψ_0, of $\mathcal{L}^* - V$ corresponding to λ_0, is that they allow averages with respect to ψ_0 to be written in terms of averages with respect to the invariant measure of $Y^{(t)}$ without any appearance of the potential V which we have already seen can lead to explosive growth of variance.

Of course in practice we do not have access to either u or φ_0 and precomputing either would be as difficult as the original sampling problem. Nonetheless, they serve as useful guides in designing effective importance sampling strategies. Our goal should always be to design reference densities for importance sampling that are as "close" to the q specified by Doob's h-transformation as possible. In some cases this can be done using less expensive (and less accurate) numerical approaches to pre-estimate u or φ_0, and in other cases one may have to rely on physical intuition.

10.6 Branching processes and diffusion Monte Carlo

In this section we will consider a very different stochastic representation of the solution to the backwards Kolmogorov equation (10.2). We will define a measure valued Markov process $\Phi^{(t)}$ with the property that

$$E_{\delta_x}\left[\Phi^{(t)}[f]\right] = E_x\left[f(X^{(t)})e^{-\int_0^t V(s,X^{(s)})ds}\right] \tag{10.17}$$

where $X^{(t)}$ solves the martingale problem (10.1) and we have introduced the notation

$$\Phi^{(t)}[f] = \int f(x)\,\Phi^{(t)}(dx).$$

Before introducing this process, we will warm up with a simpler representation in a similar spirit. Let $N^{(t)}$ be a jump Markov process taking values only in the counting numbers and with increments $\triangle_t^{t+h}N \in \{-1,0,1\}$ obeying the rule

$$P\left[\triangle_t^{t+h}N = \pm 1 \,|\, \mathcal{F}_t\right] = V_{\mp}(s, X^{(t)})\,N^{(t)}\,h + o(h)$$

where $V_+ = \max\{V, 0\}$ and $V_- = \max\{-V, 0\}$. We will also assume that $N^{(0)} = 1$ and that if $N^{(t)} = 0$ then $N^{(s)} = 0$ for all $s \geq t$. By now, we are very familiar with kind of process and we know that it can be written as

$$N^{(t)} = 1 + N_+^{(t)} - N_-^{(t)}$$

where $N_{\pm}^{(t)}$ are Poisson processes with respective intensities $V_{\mp}(s, X^{(t)})\,N^{(s)}$ that, conditioned on their intensity processes, are independent of one another. Taking the expectation of $N^{(t)}$ conditioned on the path of $X^{(t)}$ up to time t we find that

$$\frac{d}{dt}\mathbf{E}\left[N^{(t)}\,|\,\{X^{(s)}\}_{s\leq t}\right] = -V(s, X^{(t)})\,E\left[N^{(t)}\,|\,\{X^{(s)}\}_{s\leq t}\right]$$

which implies that

$$E\left[N^{(t)}\,|\,\{X^{(s)}\}_{s\leq t}\right] = e^{-\int_0^t V(s,X^{(s)})ds}.$$

From this last expression we can conclude that

$$E_x\left[f(X^{(t)})e^{-\int_0^t V(s,X^{(s)})ds}\right] = E_x\left[f(X^{(t)})N^{(t)}\right].$$

One can easily show that the expression inside the expectation on the right has higher variance than the expression inside the expectation on the left. We have introduced this seemingly unnecessary additional noise to make the point that the exponential factor in our expectation can be replaced by an integer valued random process. The key idea behind our measure valued process will be to realize that $N^{(t)}$ can be interpreted as a number of copies of the underlying process $X^{(t)}$.

The measure $\Phi^{(t)}$ will at all times be of the form

$$\Phi^{(t)} = \sum_{\ell=0}^{N^{(t)}-1} \delta_{X^{(t,\ell)}}$$

with $N^{(0)} = 1$, $X^{(0,0)} = x$, and each $X^{(t,\ell)}$ evolving independently according to $\mathcal{L}$. With $\Phi^{(t)}$ of this form,

$$\Phi^{(t)}[f] = \sum_{\ell=0}^{N^{(t)}-1} f(X^{(t,\ell)}).$$

Associated with each of the $X^{(t,\ell)}$ is a Markov jump process $N^{(t,\ell)}$ of the form

$$N^{(t,\ell)} = 1 + N_+^{(t,\ell)} - N_-^{(t,\ell)}$$

where $N_\pm^{(t,\ell)}$ are Poisson processes with respective intensities $V_\mp(t, X^{(t,\ell)})\, N^{(t,\ell)}$ that, conditioned on their intensities are independent of each other and of the $X^{(t,\ell)}$ processes. The key difference with the previous construction is that now, as soon as $N^{(t,\ell)} = 2$, a new "copy" $(X^{(t,k)}, N^{(t,k)}) = (X^{(t,\ell)}, 1)$ is introduced for $k = N^{(t)} + 1$, $N^{(t,\ell)}$ is reset to 1, and $N^{(t)}$ is incremented by 1. The two pairs $(X^{(t,\ell)}, N^{(t,\ell)})$ and $(X^{(t,k)}, N^{(t,k)})$ subsequently evolve independently. When $N^{(t,\ell)} = 0$ the copy $(X^{(t,\ell)}, N^{(t,\ell)})$ is deleted and $N^{(t)}$ is decremented by 1.

To see that (10.17) holds define the function

$$u(r, x) = E_{r,\delta_x}\left[\sum_{\ell=0}^{N^{(t)}-1} f(X^{(t,\ell)})\right]$$

for $r \leq t$ and notice that for any $h \geq 0$,

$$u(r, x) = E_{r,\delta_x}\left[\sum_{\ell=0}^{N^{(r+h)}-1} u(r+h, X^{(r+h,\ell)})\right]. \qquad (10.18)$$

The process $N^{(t)}$ can be written

$$N^{(t)} = 1 + N_+^{(t)} - N_-^{(t)}$$

where $N_\pm^{(t)}$ are Poisson processes with respective intensities

$$\sum_{\ell=0}^{N^{(t)}-1} V_\mp(t, X^{(t,\ell)}).$$

Therefore

$$P\left[\triangle_t^{t+h} N = \pm 1 \,|\, \Phi^{(t)}\right] = h \sum_{\ell=0}^{N^{(t)}-1} V_\mp(t, X^{(t,\ell)}) + o(h).$$

and

$$P\left[\triangle_t^{t+h} N = 0 \,|\, \Phi^{(t)}\right] = 1 - h \sum_{\ell=0}^{N^{(t)}-1} |V|(t, X^{(t,\ell)}) + o(h).$$

In particular, these expressions imply that when we partition the expectation on the right hand side of (10.18) into the events $N^{r+h} = 0$, $N^{(r+h)} = 1$, $N^{(r+h)} = 2$, and $N^{(r+h)} > 2$, we obtain

$$\begin{aligned} u(r,x) = E_{r,x}\left[u(r+h, X^{(r+h)})\right](1 - h|V|(r,x)) \\ + 2h\, u(r,x)\, V_-(r,x) + o(h) \end{aligned}$$

where in the $N^{(r+h)} = 2$ term we have used the facts that $E_{r,\delta_x}\left[\left(X^{(r+h,\ell)} - x\right)^2\right] = \mathcal{O}(h)$ and that $u(r+h,x) - u(r,x) = o(h)$. Expanding the first term we find that

$$\begin{aligned} u(r,x) &= u(r+h,x) + h\,\mathcal{L}u(r,x) \\ &\quad - h\,u(r,x)\,|V|(r,x) + 2h\,u(r,x)\,V_-(r,x) + o(h) \\ &= u(r+h,x) + h\,\mathcal{L}u(r,x) - hu(r,x)\,V(r,x) + o(h) \end{aligned}$$

which allows us to conclude that u solves (10.2) and, assuming that solutions to that equation are unique, establishes (10.17).

Notice that the expectation in (10.17) is only one relatively simple moment of the process $\Phi^{(t)}$ and does not completely characterize its evolution. To completely characterize the evolution of $\Phi^{(t)}$ we can compute its generator. As for many other processes, this can be done by considering the evolution of the moment generating functions of $\Phi^{(t)}$. For any initial empirical measure of the form

$$\mu = \sum_{\ell=0}^{n-1} \delta_{x_\ell}$$

and any fixed t, consider the expectations

$$\varphi_f(h) = E_{t,\mu}\left[e^{\Phi^{(t+h)}[\log f]}\right] = \prod_{\ell=0}^{n-1} E_{t,\delta_{x_\ell}}\left[\prod_{\ell=0}^{N^{(t+h)}-1} f(X^{(t+h,\ell)})\right]$$

with $f \geq 0$. By partitioning into the events $N^{(t+h)} = 0$, $N^{(t+h)} = 1$, $N^{(t+h)} = 2$, and $N^{(t+h)} > 2$, and expanding terms just as before we find that

$$\begin{aligned}
\varphi_f(h) &= \prod_{\ell=0}^{n-1} \Big(h\, V_+(t, x_\ell) + E_{t,x_\ell}\left[f(X^{(t+h)})\right](1 - h|V|(t, x_\ell)) \\
&\qquad\qquad + h\, v^2(t, x_\ell)\, V_-(t, x_\ell) + o(h)\Big) \\
&= \prod_{\ell=0}^{n-1} \Big(f(x_\ell) + h\, \mathcal{L}f(x_\ell) + h\, \big(V_+(t, x_\ell) - f(x_\ell)\, |V|(t, x_\ell) \\
&\qquad\qquad + f^2(x_\ell)\, V_-(t, x_\ell)\big)\Big) + o(h) \\
&= e^{\mu[\log f]} \mu \left[1 + h\, \frac{\mathcal{L}f + V_+ - f|V| + f^2 V_-}{f}\right] + o(h).
\end{aligned}$$

Now assembling

$$\frac{\varphi_f(h) - \varphi_f(0)}{h} = e^{\mu[\log f]} \mu \left[\frac{\mathcal{L}f + V_+ - |V|f + V_- f^2}{f}\right] + o(1)$$

we find that the action of the generator of $\Phi^{(t)}$ on moment generating functions is

$$e^{\mu[\log f]} \longmapsto e^{\mu[\log f]} \mu \left[\frac{\mathcal{L}f + V_+ - |V|f + V_- f^2}{f}\right]$$

which completely characterizes the evolution of $\Phi^{(t)}$.

10.7 bibliography