Lecture 4 - Mathematical Foundations of Stochastic Processes (substitute lecturer Oliver Bühler)

Oliver Bühler; notes by Tom Eaves & Anna FitzMaurice

June 18, 2015

1 Itō calculus

Recall that for the stochastic differential equation

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W_t \tag{1}$$

we have the important relations

$$\mathbb{E}(g(X_t)\mathrm{d}W_t) = 0 \tag{2}$$
$$\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t \tag{3}$$
$$\mathbb{E}(\mathrm{d}W(t_1)\mathrm{d}W(t_2)) = \delta(t_1 - t_2)\mathrm{d}t_1\mathrm{d}t_2 \tag{4}$$

We would like to reconcile the fact that we have terms of order $\mathrm{d}t$ and terms of order $\sqrt{\mathrm{d}t}$ in the same equation. Essentially, $\mathrm{d}W_t$ is large and incoherent, whereas $\mathrm{d}t$ is small but coherent, and they act together to result in equal contributions. The fact that $\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t$ means that when attempting to work with the chain rule when changing variables, we need to evaluate more derivatives than expected in order to complete the stochastic differental equation to the correct order. For example,

$$\mathrm{d}(F(X_t)) = F'(X_t)\mathrm{d}X_t + F''(X_t)\mathrm{d}X_t\mathrm{d}X_t/2 + o(\mathrm{d}X_t\mathrm{d}X_t). \tag{5}$$

Additionally, $\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t$ requires careful interpretation. Recall that for *finite* increments in the Weiner process,

$$\mathbb{E}(\Delta W^2) = \Delta t \tag{6}$$

and so the infinitesimal statement should be interpreted as any errors associated with approximating $\Delta W^2 \approx \Delta t$ are canceled in the limit of infinitesimal increments which are then summed over as an integral, and this process works because we are summing a family of independent Gaussian-distributed random variables.

With the Itō calculus rule (5) we may re-examine the examples already considered without reference to the Fokker–Planck equation.

1.1 Ornstein-Uhlenbeck Equation

Consider again the Ornstein-Uhlenbeck stochastic differential equation

$$\mathrm{d}U_t = -\gamma U_t \mathrm{d}t + \sigma \mathrm{d}W_t. \tag{7}$$

We can in fact integrate this exactly by multiplying through by $e^{\gamma t}$ to get

$$\mathrm{d}(e^{\gamma t} U_t) = e^{\gamma t}\sigma \mathrm{d}W_t, \tag{8}$$

and so

$$U_t = U_0 e^{-\gamma t} + \sigma e^{-\gamma t}\int_0^t e^{\gamma s}\mathrm{d}W_s, \tag{9}$$

which gives the expectation

$$\mathbb{E}(U_t) = e^{-\gamma t}\mathbb{E}(U_0) \tag{10}$$

and variance

$$\mathbb{E}(U_t^2) = \sigma^2 e^{-2\gamma t}\int_0^t\int_0^t e^{\gamma(s_1+s_2)}\mathbb{E}(\mathrm{d}W(s_1)\mathrm{d}W(s_2)). \tag{11}$$

We could instead obtain these results directly from the stochastic differential equation by forming an equation for $\mathrm{d}(U_t^2)$ using Itō calculus. From the Itō formula (5) we have

$$U_t \mathrm{d}U_t = \frac{\mathrm{d}(U_t^2)}{2} - \frac{\mathrm{d}U_t \mathrm{d}U_t}{2}, \tag{12}$$

and from (7) and the relation $\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t$,

$$\mathrm{d}U_t\mathrm{d}U_t = \sigma^2\mathrm{d}t + o(\mathrm{d}t), \tag{13}$$

which gives

$$\frac{\mathrm{d}(U_t^2)}{2} - \frac{\sigma^2}{2}\mathrm{d}t = -\gamma U_t^2\mathrm{d}t + U_t\sigma\mathrm{d}W_t. \tag{14}$$

In steady state $\mathbb{E}^s(\mathrm{d}) = 0$ and so

$$\frac{\sigma^2}{2} = \mathbb{E}^s(\gamma U_t^2), \tag{15}$$

which is the fluctuation-dissipation relation for this process.

1.2 Linear population model

Consider the stochastic differential equation

$$\mathrm{d}X_t = \mu X_t\mathrm{d}t + \sigma X_t\mathrm{d}W_t. \tag{16}$$

This equation can be interpreted as a random interest rate model.

We have

$$\mathbb{E}(\mathrm{d}X_t) = \mu\mathbb{E}(X_t\mathrm{d}t), \tag{17}$$

and so

$$\mathbb{E}(X_t) = X_0 e^{\mu t}. \tag{18}$$

We also have

$$\mathrm{d}(X_t^n) = nX_t^{n-1}\mathrm{d}X_t + \frac{n(n-1)X_t^{n-2}\mathrm{d}X_t\mathrm{d}X_t}{2} \tag{19}$$

$$= \left(n\mu + \frac{n(n-1)\sigma^2}{2}\right) X_t^n\mathrm{d}t + n\sigma X_t^n \mathrm{d}W_t, \tag{20}$$

and so the n-th moment is

$$\mathbb{E}(X_t^n) = X_0^n \exp\left[\left(n\mu + \frac{n(n-1)\sigma^2}{2}\right) t\right]. \tag{21}$$

Alternatively, using the Itō calculus formula (5), we note that in the absence of noise, we would be interested in $\mathrm{d}(\log(X))$, and so we compute

$$\mathrm{d}(\log(X_t)) = \frac{\mathrm{d}X_t}{X_t} - \frac{\mathrm{d}X_t\mathrm{d}X_t}{2X_t^2} = \frac{\mathrm{d}X_t}{X_t} - \frac{\sigma^2\mathrm{d}t}{2} \tag{22}$$

$$= \mu\mathrm{d}t - \frac{\sigma^2}{2}\mathrm{d}t + \sigma\mathrm{d}W_t. \tag{23}$$

Hence, the solution is

$$X_t = X_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right]. \tag{24}$$

We can reconcile the fact that at first glance the results $\mathbb{E}(W_t) = 0$ and $\mathbb{E}(X_t) = X_0 e^{\mu t}$ appear incompatible with this solution by recognising that the occasions for which $W_t > 0$ and $W_t < 0$ do not contribute equally after exponentiating.

From this solution we see that if $\sigma^2 > 2\mu$, then extinction is guaranteed almost surely, as for the nonlinear population model discussed in a previous lecture. This is since near extinction, X is small, and so the linearised approximation is accurate.

We can find the transition density $\rho(x, t|x_0, 0)$ by solving the Fokker–Planck equation for $\rho_Y(y, t|y_0, 0)$ for the variable $Y_t = \log(X_t)$, since

$$\frac{\partial \rho_Y}{\partial t} = \left(\frac{\sigma^2}{2} - \mu\right)\frac{\partial \rho_Y}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2 \rho_Y}{\partial y^2}, \tag{25}$$

which can be solved with Fourier transforms to give

$$\rho_Y = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{\left(y - y_0 - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}\right], \tag{26}$$

and so Y_t is normally distributed, meaning that X_t is log-normally distributed, with

$$\rho(x, t|x_0, 0) = \frac{1}{x\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{\left(\log x - \log x_0 - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}\right]. \tag{27}$$

Now let $\epsilon > 0$. Then,

$$\mathbb{P}(X_t > \epsilon | X_0 = x_0) = \int_{\epsilon}^{\infty} \rho(x, t|x_0, 0)\mathrm{d}x = \frac{1}{\sqrt{\pi}}\mathrm{erfc}\left(\frac{\log \epsilon - \log x_0 - (\mu - \frac{\sigma^2}{2})t}{\sqrt{2\sigma^2 t}}\right), \quad (28)$$

and so if $\sigma^2 > 2\mu$, we have that $\mathbb{P}(X_t > \epsilon | X_0 = x_0) \to 0$ as $t \to \infty$, despite the fact that $\mathbb{E}(X_t) = X_0 \exp(\mu t)$. Almost all trajectories decay eventually, but the moments of the distribution grow rapidly, and so in an ensemble we expect an occasional 'success'. We can see that the distribution becomes more shifted towards $x = 0$ as time t increases in Figure 1, in which the distribution for $\mu = 1$, $\sigma = 2$ and $x_0 = 1$ is plotted at various times.

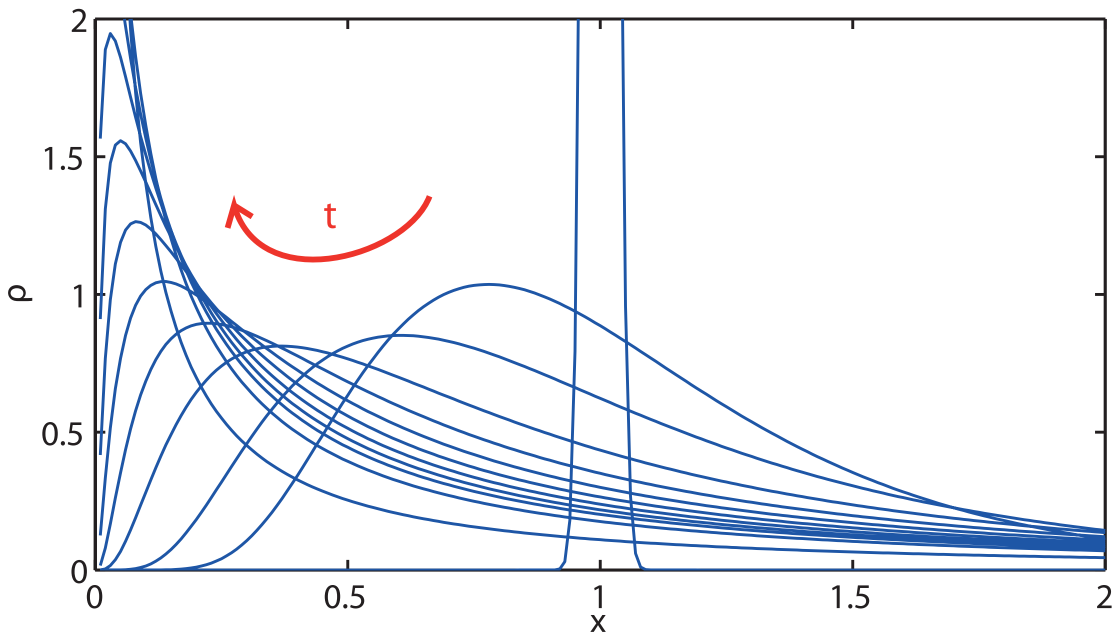

Figure 1: The transition density $\rho(x, t|1, 0)$ for the linear population growth model with $\mu = 1$ and $\sigma = 2$ at times $t = 0.0001$, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1 and 2.

1.3 Nonlinear population model

Consider the stochastic differential equation

$$\mathrm{d}X_t = (\mu X_t - X_t^2)\mathrm{d}t + \sigma X_t \mathrm{d}W_t. \quad (29)$$

Let $Y_t = X_t^{-1}$. Then, the Itō calculus formula (5) gives

$$\mathrm{d}Y_t = -\frac{\mathrm{d}X_t}{X_t^2} + \frac{\mathrm{d}X_t \mathrm{d}X_t}{X_t^3} \quad (30)$$

$$= -\frac{(\mu X_t - X_t^2)\mathrm{d}t + \sigma X_t \mathrm{d}W_t}{X_t^2} + \frac{\sigma^2 X_t^2 \mathrm{d}t}{X_t^3} \quad (31)$$

$$= (1 - (\mu - \sigma^2)Y_t)\mathrm{d}t - \sigma Y_t \mathrm{d}W_t, \quad (32)$$

which is linear in Y_t.

Then,

$$\mathbb{E}(\mathrm{d}Y_t) = -(\mu - \sigma^2)\mathbb{E}(Y_t \mathrm{d}t), \tag{33}$$

which gives

$$\mathbb{E}(Y_t) = Y_0 e^{-(\mu-\sigma^2)t} \to \infty \text{ as } t \to \infty \text{ if } \mu < \sigma^2, \tag{34}$$

as expected from a previous lecture.

2 Probability currents and steady states

Recall the Fokker-Planck equation for $\mathbf{x} \in \mathbf{R}^n$ and noise $\mathbf{w} \in \mathbf{R}^m$ subject to

$$\mathrm{d}X_i = f_i \mathrm{d}t + g_{ij}\mathrm{d}W_j, \tag{35}$$

namely

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{f}) = \frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{D_{ij}}{2}\rho\right), \tag{36}$$

where $D_{ij} = g_{ik}g_{jk}$ is an $n \times n$ matrix.

We can introduce a *probability current* $\mathbf{J}$ by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{37}$$

and so

$$J_i = f_i \rho - \frac{\partial}{\partial x_j}\left(\frac{D_{ij}}{2}\rho\right) \tag{38}$$

To have a steady state we need $\nabla \cdot \mathbf{J} = 0$, which can be achieved in two ways.

1. $\mathbf{J} = 0$ corresponds to equilibrium solutions, or detailed balance solutions in which each point of any boundary has zero net flux across it.

2. $\mathbf{J} \neq 0$ corresponds to solutions with flux.

First consider $\mathbf{J} = 0$. Then, write $\rho = e^{-\phi} > 0$. The condition $\mathbf{J} = 0$ becomes

$$\frac{D_{ij}}{2}\frac{\partial \phi}{\partial x_j} = -(f_i + v_i), \tag{39}$$

where

$$v_i = -\frac{\partial}{\partial x_j}\left(\frac{D_{ij}}{2}\right), \tag{40}$$

and so provided that D_{ij} is invertible the solution is obtained from

$$\frac{\partial \phi}{\partial x_j} = -2D_{ij}^{-1}(v_i + f_i). \tag{41}$$

Given that the left hand side of this equation is $\nabla\phi$, we have a compatibility condition for the existence of such a solution,

$$\mathrm{curl}(\mathbf{D}^{-1} \cdot (\mathbf{v} + \mathbf{f})) = 0. \tag{42}$$

We now show some examples of density currents satisfying $\nabla \cdot \mathbf{J} = 0$.

2.1 Uniform noise

This is an example of $\nabla \cdot \mathbf{J} = 0$ achieved through $\mathbf{J} = 0$. Let $D_{ij} \propto \delta_{ij}$. Then (41) becomes $\mathbf{f} = \nabla\phi$, which is a *gradient drift* solution, and for compatibility we require $\nabla \times \mathbf{f} = 0$.

2.2 Gradient flow plus Hamiltonian flow

In two dimensions, let $\mathbf{f} = -\sigma^2\nabla\phi + \nabla \times (H\hat{\mathbf{x}}_3)$ and $D_{ij} = 2\sigma^2\delta_{ij}$.

The corresponding deterministic equation would be

$$\dot{x_1} = -\phi_{x_1} - H_{x_2}, \tag{43}$$

$$\dot{x_2} = -\phi_{x_2} + H_{x_1}, \tag{44}$$

i.e. the sum of a gradient flow ϕ and a Hamiltonian flow H.

An example of such a stochastic differential equation would be noisy rotating decay

$$\mathrm{d}U_t = -\gamma U_t\mathrm{d}t + fV_t\mathrm{d}t + \sigma\mathrm{d}W_{t_1}, \tag{45}$$

$$\mathrm{d}V_t = -\gamma V_t\mathrm{d}t - fU_t\mathrm{d}t + \sigma\mathrm{d}W_{t_2}. \tag{46}$$

For this $\mathbf{f}$ and $\mathbf{D}$, try the solution $\rho = e^{-\phi}$ to get

$$\mathbf{J} = e^{-\phi}\mathbf{f} + \sigma e^{-\phi}\nabla\phi. \tag{47}$$

When taking the divergence, many terms cancel, and we are left with

$$\nabla \cdot \mathbf{J} = -e^{-\phi}\nabla\phi \cdot \nabla \times (H\hat{\mathbf{x}}_3), \tag{48}$$

and so we obtain $\nabla \cdot \mathbf{J} = 0$ provided that

$$J(\phi, H) = 0, \tag{49}$$

where $J(\cdot\,,\cdot)$ is the Jacobian.

2.3 Forced harmonic oscillator

The stochastic differential equation

$$\mathrm{d}X_t = Y_t\mathrm{d}t, \tag{50}$$

$$\mathrm{d}Y_t = -X_t\mathrm{d}t - \gamma Y_t\mathrm{d}t + \sigma\mathrm{d}W_t, \tag{51}$$

has stationary solution

$$\rho^s = \mathcal{N}\exp\left(-\frac{\gamma}{\sigma^2}(x^2 + y^2)\right). \tag{52}$$

3 Kolmogorov Backward Equation

We have so far considered the Fokker-Planck equation, which tells us the evolution forwards in time of a probability distribution for a given SDE from a corresponding initial condition. We now derive the Kolmogorov Backward Equation (KBE), which can be thought of as the PDE governing the evolution of a distribution backwards in time, and will subsequently demonstrate the application of this equation to a variety of problems of interest.

3.1 Derivation of the KBE

Starting from the Chapman-Kolmogorov equation (which follows simply from the Markov property for all Markovian processes)

$$\rho(x,t|y,s) = \int_{-\infty}^{\infty} \rho(x,t|x',t')\rho(x',t'|y,s)\ \mathrm{d}x', \tag{53}$$

differentiate with respect to t' to obtain

$$0 = \int_{-\infty}^{\infty} \left[\frac{\partial\rho}{\partial t'}(x,t|x',t')\rho(x',t'|y,s) + \rho(x,t|x',t')\frac{\partial\rho}{\partial t'}(x',t'|y,s)\right]\ \mathrm{d}x'. \tag{54}$$

Now substituting for $\frac{\partial\rho}{\partial t'}(x',t'|y,s)$ using the Fokker-Planck (forward) equation, and using integration by parts

$$\begin{aligned}
0 = \int_{-\infty}^{\infty} \mathrm{d}x'\ \ & \frac{\partial\rho}{\partial t'}(x,t|x',t')\rho(x',t'|y,s) \\
& + \rho(x,t|x',t')\frac{\partial}{\partial x'}\left(-f(x') + \frac{1}{2}\frac{\partial}{\partial x'}g(x')^2\right)\rho(x',t'|y,s) \\
0 = \int_{-\infty}^{\infty} \mathrm{d}x'\ \ & \rho(x',t'|y,s)\left[\frac{\partial\rho}{\partial t'}(x,t|x',t') + f(x')\frac{\partial\rho}{\partial x'}(x,t|x',t') + \frac{1}{2}g(x')^2\frac{\partial^2\rho}{\partial x'^2}(x,t|x',t')\right].
\end{aligned}$$

If we now let the time interval $|t'-s| \to 0$, then $\rho(x',t'|y,s) \to \delta(x'-y)$, so we are left with

$$0 = \frac{\partial\rho}{\partial s}(x,t|y,s) + f(y)\frac{\partial\rho}{\partial y}(x,t|y,s) + \frac{1}{2}g(y)^2\frac{\partial^2\rho}{\partial y^2}(x,t|y,s), \tag{55}$$

or

$$\frac{\partial\rho}{\partial s}(x,t|y,s) = \left(f(y)\frac{\partial}{\partial y} + \frac{1}{2}g(y)^2\frac{\partial^2}{\partial y^2}\right)\rho(x,t|y,s), \tag{56}$$

which is known as the Kolmogorov Backward Equation. Note that the operator $\mathcal{L} \equiv f\partial_x + \frac{g^2}{2}\partial_x^2$ is the formal adjoint of the forward Fokker-Planck operator $\mathcal{L}^\dagger \equiv \partial_x(-f + \partial_x\frac{g^2}{2})$.

3.2 Survival times and first passage times

The power of the KBE becomes transparent if we consider the problem of a random process on some specified domain, and wish to make statements about the time taken for the process to stray outside the domain (variously known as the first passage time, the exit time, the escape time, the stopping time, or the hitting time of the process), or if we wish to determine the region of the boundary through which the process exits the domain.

3.2.1 Survival time

Consider a 1D process $X(t)$ on $x \in (x_a, x_b)$, and impose absorbing boundary conditions $\rho(x_a,t|x_0,t_0) = \rho(x_b,t|x_0,t_0) = 0$. It is of interest to compute the *survival probability* $S(t|x_0,t_0) \equiv \mathbb{P}(x_a < X(u) < x_b\ \forall u < t)$. By definition, S is a monotonically decreasing function of t, with $S(t_0|x_0,t_0) = 1$ for $x_0 \in (x_a,x_b)$ and $S(t|x_0,t_0) \to 0$ as $t \to \infty$, so probability can be thought of as "leaking" over the edge of the domain as time progresses. Thus probability density is not conserved, and it can be seen that $S(t|x_0, t_0)$ is given by

$$S(t|x_0, t_0) = \int_{x_a}^{x_b} \rho(x, t|x_0, t_0) \, \mathrm{d}x. \tag{57}$$

Consequently we may obtain a PDE for S by integrating the KBE for the process, as follows

$$\int_{x_a}^{x_b} \left[\frac{\partial \rho}{\partial t_0}(x, t|x_0, t_0) = \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) \rho(x, t|x_0, t_0) \right] \, \mathrm{d}x \tag{58}$$

$$-\frac{\partial}{\partial t_0} S(t|x_0, t_0) = \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) S(t|x_0, t_0), \tag{59}$$

which can be solved for survival time S given boundary conditions $S(t|x_a, t_0) = S(t|x_b, t_0) = 0$ and initial condition $S(t_0|x_0, t_0) = 1$ for $x_a < x_0 < x_b$.

3.2.2 First passage time

For the 1D process above, define random variable t_{exit} as the first time at which $X = x_a$ or $X = x_b$. Then the mean exit time for a process starting at (x_0, t_0) is, by definition

$$\mathbb{E}_{x_0}(t_{\text{exit}} - t_0) = \int_{t_0}^{\infty} (t - t_0) p(t|x_0, t_0) \, \mathrm{d}t, \tag{60}$$

where $p(t|x_0, t_0) \equiv -\frac{\mathrm{d}}{\mathrm{d}t} S(t|x_0, t_0)$ is the probability of absorption at time t. Integrating by parts,

$$\mathbb{E}_{x_0}(t_{\text{exit}} - t_0) = -\Big[(t - t_0) S(t|x_0, t_0) \Big]_{t_0}^{\infty} + \int_{t_0}^{\infty} S(t|x_0, t_0) \, \mathrm{d}t, \tag{61}$$

and it can be seen that the boundary terms vanish provided $S(t|x_0, t_0) \sim o(t^{-1})$ as $t \to \infty$, which holds provided the mean survival time is well-defined, so

$$\mathbb{E}_{x_0}(t_{\text{exit}} - t_0) = \int_{t_0}^{\infty} S(t|x_0, t_0) \, \mathrm{d}t. \tag{62}$$

To get an equation for the mean exit time, we then integrate equation (59) between (t_0, ∞) to give

$$-\frac{\partial}{\partial t_0} \int_{t_0}^{\infty} S(t|x_0, t_0) \, \mathrm{d}t - S(t_0|x_0, t_0) = \int_{t_0}^{\infty} \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) S(t|x_0, t_0) \, \mathrm{d}t, \tag{63}$$

using Leibniz's rule. Now noting that the mean exit time is independent of t_0 for an autonomous system, and that $S(t_0|x_0, t_0) = 1$, we have

$$-1 = \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) \mathbb{E}_{x_0}(t_{\text{exit}} - t_0), \tag{64}$$

which may be solved for mean exit time $\mathbb{E}_{x_0}(t_{\text{exit}} - t_0)$.

3.3 Alternative derivation: Change of variables

An alternative derivation of the KBE to that above is to consider the change of variables $Y = u(X,t)$ in the SDE $\mathrm{d}X = f(X)\mathrm{d}t + g(X)\mathrm{d}W$, for some function u. Using Itō calculus, the change of variables becomes

$$\mathrm{d}Y = u_t \mathrm{d}t + u_X \mathrm{d}X + \frac{1}{2} u_{XX} \mathrm{d}X \mathrm{d}X \tag{65}$$

$$= (u_t + f u_X + \frac{g^2}{2} u_{XX})\mathrm{d}t + u_X g \mathrm{d}W, \tag{66}$$

on substituting for $\mathrm{d}X$ using the governing SDE, and noting that $\mathrm{d}W\mathrm{d}W = \mathrm{d}t$. We then have

$$\mathrm{d}Y = (u_t + \mathcal{L}u)\mathrm{d}t + u_X g \mathrm{d}W \tag{67}$$

for operator $\mathcal{L} \equiv f\partial_x + \frac{g^2}{2}\partial_x^2$, as before. The KBE is precisely the equation $u_t + \mathcal{L}u = 0$, and (as a backward heat equation) is well-posed when conditions are specified on some final time $t = T > t_0$. It can be seen from integrating equation (67) that the solution to the homogeneous problem $u_t + \mathcal{L}u = 0$ with condition $u(X,T) = \phi(X)$ generates the expectation $u(X,T) = \mathbb{E}(\phi(X(T))|X(t) = x)$, so for this reason $\mathcal{L}$ is sometimes referred to as the *generator*.