A stochastic differential equation is a particular type of noisy dynamical system. Applications call for generating SDE sample paths subject to a variety of constraints.

1 Multivariate normal

Let $X \in R^n$ be an n component normal with covariance matrix C. Choose an orthonormal family of vectors v_k and eigenvalues λ_k so that $Cv_k = \lambda_k v_k$. The λ_k are the *principle components* of X (or of C) and the v_k the corresponding *principle components* (the terminology differs slightly from place to place). The random numbers $Y_k = v_k^t X$ are independent mean zero Gaussians with mean zero and variance λ_j. Said another way, if V is the orthogonal matrix whose columns are the v_k, then $Y = V^t X$ is a multivariate normal with mean zero and covariance Λ, the diagonal matrix with λ_k in the (k,k) position. The Y_k also may be called principle components, or possibly components in the principle directions. To summarize, the eigenvalue/eigenvector decomposition of C is the *principle component analysis* of X. When using PCE, we usually order the principle values from largest to smallest: $\lambda_k \geq \lambda_{k+1}$ for all k.

On the other hand, suppose we know the v_k and λ_k. We can construct a sample Y_k as $Y_k = \sigma_k Z_k$, where $\sigma_k = \sqrt{\lambda_k}$ and the Z_k are independent standard normals. Therefore, the formula $X = \sum_{k=1}^n \sigma_k Z_k$ provides a sample of X from independent standard normals. This kind of PCA is helpful in many ways in Monte Carlo practice. If we want to know $A = E[f(X)]$, it is likely that the larger principle components are more important than the smaller ones. For example, the quantity $B = \int f(\sigma_1 v_1 z_1) e^{-z_1^2/2} dz/\sqrt{2\pi}$ may be a helpful control variate for evaluating A.

2 Brownian motion

A standard univariate Brownian motion path is a continuous random function $X(t)$ with $X(0) = 0$ and *increments*, $\Delta X_k = X(t_{k+1}) - X(t_k)$ independent Gaussian random variables with mean zero and variance $t_{k+1} - t_k$. Here t_k is any sequence of positive numbers with $t_{k+1} > t_k$. Many problems involve expected values of functions of Brownian motion and the intuitions gained by studying Brownian motion help design methods for related problems.

We start with the PCA of standard Brownian motion. Since Brownian motion is infinite dimensional, we start with finite dimensional approximations then get the PCA of Brownian motion itself as a limit. Suppose we fix a $T > 0$. For any integer $n > 0$, we divide the interval $[0, T]$ into n equal subintervals using the notation $h = T/n$ and $t_k = kh$. The multivariate normal $X^h \in R^n$ has components $X_k^h = X(t_k)$, for $1 \leq k \leq n$. We showed before that the covariance matrix for X^h has $H = C^{-1}$ that is tri-diagonal of the form .... . Clearly the eigenvalues of C are the inverses of the eigenvalues of H while the eigenvectors are the same. It happens (we will get more insight into this later)

that the eigenvectors of H have the form $v^h_{k,j} = \sin(\omega^h_k t^h_j)$, where ... . If we take the limit $h \to 0$ and $n \to \infty$ with $T = nh$ fixed, the sequence of eigenvalues of H converges to $\mu_k = (k+1)^2\pi^2/T^2$, as we verify by direct computation.

The PCA representation of a Brownian motion path is

$$X(t) = \frac{T}{\pi} \sum_{k=1}^{\infty} \frac{Z_k}{k+\frac{1}{2}} \sin\left(\frac{k+\frac{1}{2})\pi}{T}\, t\right) \quad . \tag{1}$$

This representation is valid only for t in the range $0 \le t \le T$. Moreover, it presents serious mathematical challenges. It is not straightforward to see that the sum converges. Indeed, the Z_k and the sine factors are of the order of unity, so the terms in the sum seem to have the order of magnitude $1/k$, which would not by itself lead to a convergent sum. The series does converge in the L^2 sense because

$$E\left[\sum_{k=1}^{\infty} \frac{Z_k^2}{\left(k+\frac{1}{2}\right)^2}\right] \;=\; \sum_{k=1}^{\infty} \frac{1}{\left(k+\frac{1}{2}\right)^2} \;<\; \infty \,.$$

However, this kind of convergence does not even imply that $X(t)$ is continuous. Many discontinuous are represented by L^2 convergence Fourier series. If we wish to generate a Brownian motion path using (**??**), we can use the FFT to calculate the sums on the right efficiently.

The Brownian Bridge construction is another good way to generate Brownian motion paths. For this, we consider a *diadic* collection of times. On level k, the interval $[0,T]$ is divided into 2^k equal subintervals. At the top is level zero with a single interval $[0,T]$ whose endpoints may be written $t_{0,0} = 0$ and $t_{0,1} = T$. At the next level $[0,T]$ is broken into two halves bounded by the points $t_{1,0} = 0$, $t_{1,1} = T/2$, and $t_{1,2} = T$. The diadic points at level k are $t_{k,j} = j2^{-k}T$, for integers j in the range $0 \le j \le 2^k$.

The construction itself is an algorithm that generates values for $X_{k,j} = X(t_{k,j})$ recursively proceeding from k to $k+1$. At the top level we already have $X_{0,0} = X(0) = 0$ and we need $X_{0,1} = X(T)$. But $X(T)$ is a mean zero Gaussian with variance T, so we can take $X_{0,1} = \sqrt{T}Z_{0,1}$, where $Z_{0l,1}$ is a standard normal. Things get interesting at the next level. From the properties of Brownian motion, we know that the values $(X(t_{1,1}), X(t_{1,2})) = (X_{1,1}, X_{1,2})$ is a bivariate normal. Therefore, conditional on the value of $X_{1,2}$, which already is known from the previous level, the value of $X_{1,1}$ is a univariate normal. Simple calculations show that $E[X_{1,1} \mid X_{1,2}] = \frac{1}{2}X_{1,2}$, and $\mathrm{var}[X_{1,1} \mid X_{1,2}] = \frac{1}{4}T$. The first result is very natural, that the expected value of the midpoint value should be the average of the two end values. The second result may be surprising, since the variance of $X_{1,1}$ without information about $X_{1,2}$ is $T/2$ instead of $T/4$. Specifying both endpoint values constrains $X_{1,1}$ to have a smaller variance. With this information the sampling can be done using $X_{1,1} = \frac{1}{2}X_{0,1} + \frac{\sqrt{t}}{2}Z_{1,1}$. As usual, $Z_{1,1}$ is a standard normal independent of all others.

The step of going from level $k-1$ to level k is clear. We already have values $X_{k,2j} = X_{k-1,j}$ from the previous levels. We need to generate values for the new points $t_{k,2j+1}$. The Markov property (which is the main thing that makes this

construction work) implies that the conditional distribution of $X_{k,2j+1}$ given all the values on level $k-1$ is the same as the distribution given the level $k-1$ endpoints $X_{k,2j} = X_{k-1,j}$, and $X_{k,2j+2} = X_{k-1,j+1}$. As above, the conditional expectation of $X_{k,2j+1}$ is $\frac{1}{2}(X_{k-1,j} + X_{k-1,j+1})$. Let $\Delta t_k = 2^{-k}T$ be the length of the diadic intervals at level k. The conditional variance of $X_{k,2j+1}$ is $\Delta t_k/2$ (which again is half the conditional variance if only the left value $X_{k,2j}$ were specified). Therefore, the sampling may be done by

$$X_{k-2j+1} = \frac{1}{2}(X_{k-1,j} + X_{k-1,j+1}) \; + \; \sqrt{\frac{\Delta t_k}{2}} Z_{k,2j+1} \; .$$

One use of this Brownian bridge construction is similar to the PCA construction, to have a way to separate random variables that have a large impact on X from others that have a smaller impact.

3 The Ornstein Uhlenbeck Process

The *Ornstein Uhlenbeck* process is a simple modification of Brownian motion to add *mean reversion*. Starting here, we refer to the proces of interest as $X(t)$ while $W(t)$ will be Brownian motion. The general Ornstein Uhlenbeck process satisfies the SDE

$$dX = -\mu X dt + \sigma dW \; . \tag{2}$$

This means that we can write

$$X(T) = e^{-\mu T} X(0) + \sigma \int_0^T e^{-\mu(T-t)} dW(t) \; .$$

We see from this that X is a linear transformation of the Gaussian W so it is Gaussian. This is contingent on $X(0)$ being Gaussian if it is random. From this we can calculate that $X(T)$ is a gaussian with mean $e^{-\mu T}X(0)$, and variance

$$\sigma^2 \int_0^T e^{-2\mu(T-t)} dt \; = \; \frac{\sigma^2}{2\mu} \left(1 - e^{2\mu T}\right) \; \rightarrow \; \frac{\sigma^2}{2\mu} \text{ as } T \rightarrow \infty \; .$$

As $T \rightarrow \infty$, the distribution of $X(T)$ converges to a gaussian that is independent of the initial condition (if the mean and variance converge and if it's Gaussian, the distribution converges).

Much Monte Carlo depends on time dependent random processes like this so we analyze this in more detail, focusing on features that have analogues in many Monte Carlo applications. Suppose $u(x,t)$ is the probability density function for $X(t)$. This satisfies the Kolmogorov forward equation (other names include, but are not limited to, the forward equation, the Fokker Planck equation, and the Chapman Kolmogorov equation)

$$\partial_t u \; = \; \frac{\sigma^2}{2} \partial_x^2 u + \mu \partial_x (xu) \; . \tag{3}$$

We write this as

$$\partial_t u = L^* u \ , \quad \text{where } L^* u \ = \ \frac{\sigma^2}{2}\partial_x^2 u + \mu \partial_x \left(x u \right) \ . \tag{4}$$

It is traditional to call the operator in (**??**) L^* so that it's adjoint that appears in the backward equation can be called L.

If the distribution of $X(t)$ has a limit as $t \to \infty$, then we expect $u(x,t) \to u_0(x)$ as $t \to \infty$. From the equation (**??**), this suggests that u_0 is a solution with $L^* u_0 = 0$. This gives

$$\frac{\sigma^2}{2}\partial_x^2 u + \mu \partial_x \left(x u \right) = 0 \ ,$$

If $u_0 \to 0$ as $x \to \infty$ together with its derivatives, this may be integrated to

$$\frac{\sigma^2}{2}\partial_x u_0 + \mu x u_0 = 0 \ .$$

The solution takes the form $u_0(x) = Ce^{-x^2/2\rho}$, with $\rho = \sigma^2/2\mu$.

We are interested in the rate of convergence $u(\cdot,t) \to u_0$,which depends on the other eigenvalues of the operator L^*. That is, we seek functions

$$L^* u_k \ = \ \frac{\sigma^2}{2}\partial_x^2 u_k(x) + \mu \partial_x \left(x u_k \right) \ = \ -\lambda_k u_k \ . \tag{5}$$

It turns out that the solutions have the form $u_k(x) = \partial_x^k u_0$. We verify this by induction on k, the case $k = 0$ having been done above. The trick for doing the calculation is the "differentiation by parts" formula $f\partial_x g = \partial_x(fg) - (\partial_x f)g$. If $f = \partial_x^{k-1} u_0$ and $g = x$, the result is

$$x\partial_x^k u_0 \ = \ \partial_x(x\partial_x^{k-1} u_0) - \partial_x^{k-1} u_0 \ .$$

Therefore, using $L^* u_{k-1} = -\lambda_{k-1} u_{k-1}$,

$$\begin{aligned} L^* u_k \ = \ L^* \partial_x u_{k-1} \quad &= \quad \frac{\sigma^2}{2}\partial_x^2 \partial_x u_{k-1} + \mu \partial_x (x \partial_x u_{k-1}) \\ &= \quad \partial_x \left(\frac{\sigma^2}{2}\partial_x^2 u_{k-1} \right) \ + \ \partial_x \left(\ \mu \partial_x \left(x u_{k-1} \right) \ \right) \ - \ \mu \partial_x u_{k-1} \\ &= \quad \partial_x \left(L^* u_{k-1} \right) - \mu u_k \quad = \quad -\left(\lambda_{k-1} + \mu \right) u_k \ . \end{aligned}$$

This shows that $L^* u_k = -\lambda_k u_k$ with $\lambda_k = \lambda_{k-1} + \mu$. To conclude, the eigenvalues are 0 (the steady state), $-\mu$, -2μ, etc. The corresponding eigenfunctions are derivatives of Gaussians, which is to say, given by Hermite polynomials (more on this below).

We use the above information to describe the behavior of the probability density $u(x,t)$. The solution of the evolution equation (**??**) is a linear combination of eigenfunctions:

$$u(x,t) = \sum_{k=0}^{\infty} a_k e^{-k\mu t} u_k(x) \ , \tag{6}$$

where the coefficients are determined by the initial data $u(x,0)$. The $k=0$ term corresponds to the steady state solution $a_0 u_0(x)$. All the other terms decay with various exponential rates. Slowest to decay is the $k=1$ term, which will make up most of the difference between $u(x,t)$ and $a_0 u_0(x)$ for large t:

$$u(x,t) - a_0 u_0(x) \approx a_1 e^{-\mu t} u_1(x) \ , \quad \text{for large } t.$$

It may not be surprising that the exponential decay rate μ in (**??**) results in the same exponential decay rate for $u(x,t) - a_0 u_0(x)$. What is new in the PDE solution is the family of faster decay rates 2μ, 3μ, etc. We will find versions of this phenomenon in any linear Gaussian process.

There are many practical applications for this kind of analysis. One concerns dynamic sampling methods, also called Markov Chain Monte Carlo. A complicated probability distribution may not have an efficient simple sampler, but there may be a dynamic process, a Markov Chain, that preservers the distribution. Here, the dynamical process (**??**) preserves the Gaussian $Ce^{-x^2/2\rho}$. A major issue for dynamic samplers is the rate of convergence: How long do you have to run the process before the distribution of $X(t)$ closely approximates $u_0(x)$. In this simple problem, the convergence time is of the order of $1/\mu$, which also is the decay rate for the linear Gaussian process (**??**).

3.1 Hermite polynomials

Hermite polynomials have many applications in Monte Carlo and other parts of applied probability. The n^{th} *Hermite polynomial*, $H_n(x)$, is defined by

$$H_n(x) e^{-x^2/2} \ = \ \pm \partial_x^n e^{-x^2/2} \ . \tag{7}$$

The calculations $\partial_x e^{-x^2/2} = -x e^{-x^2/2}$ and $\partial_x^2 e^{-x^2/2} = (x^2-1) e^{-x^2/2}$, and $\partial_x^3 e^{-x^2/2} = (-x^3 - 3x) e^{-x^2/2}$ imply that the first few polynomials are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$ and $H_3(x) = x^3 - 3x$. It is common to choose the sign so that the leading term is positive and also to choose the sign itself always to be positive (which would give $H_1(x) = -x$). It is easy to verify by induction that (**??**) defines polynomials of degree n with leading coefficient equal to ± 1.

The H_n have the *orthogonality* property that

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-x^2/2} dx \ = \ 0 \ , \quad \text{for } n \neq m. \tag{8}$$

We briefly indicate a proof of this by induction on m, assuming $m < n$, since a better proof will come shortly. Actually, we prove (**??**) holds for a given m and all $n > m$ by induction on m. It clearly is true for $m = 0$, because then it just is $\int \partial_x^n e^{-x^2/2} dx = 0$. Now assume it is true for $m-1$ and try to prove it for m. The rewrite of (**??**) as $H_m(x) = e^{x^2/2} \partial_x^m e^{-x^2/2}$ puts (**??**) in the form

$$\int_{-\infty}^{\infty} e^{x^2/2} \left(\partial_x^n e^{-x^2/2} \right) \left(\partial_x^m e^{-x^2/2} \right) dx \ = \ 0$$

Write $\partial_x^m = \partial_x \partial_x^{m-1}$ and integrate by parts, and we get the two terms

$$-\int_{-\infty}^{\infty} e^{x^2/2} \left(\partial_x^{n+1} e^{-x^2/2}\right) \left(\partial_x^{m-1} e^{-x^2/2}\right) dx$$
$$- \int_{-\infty}^{\infty} e^{x^2/2} \left(x\partial_x^n e^{-x^2/2}\right) \left(\partial_x^{m-1} e^{-x^2/2}\right) dx$$

The top term is zero by the induction hypothesis, which is why we used a more complex induction hypothesis. We study the bottom term using something similar to the differentiation by parts idea above, namely the formula $x\partial_x^n f = \partial_x^n (xf) - n\partial_x^{n-1} f$. This implies that the bottom term is

$$\int_{-\infty}^{\infty} e^{x^2/2} \partial_x^n \left(xe^{-x^2/2}\right) \left(\partial_x^{m-1} e^{-x^2/2}\right) dx$$
$$- n\int_{-\infty}^{\infty} e^{x^2/2} \left(\partial_x^{n-1} e^{-x^2/2}\right) \left(\partial_x^{m-1} e^{-x^2/2}\right) dx$$

The bottom term is zero because $m - 1 < n - 1$. The top term is zero because $xe^{-x^2/2} = -\partial_x e^{-x^2/2}$, so it is equal to

$$-\int_{-\infty}^{\infty} e^{x^2/2} \left(\partial_x^{n+1} e^{-x^2/2}\right) \left(\partial_x^{m-1} e^{-x^2/2}\right) dx \;=\; 0\,,$$

again by induction.

The orthogonality property of Hermite polynomials gives a different way to find them. This is because of the fact that if $p(x)$ is a polynomial of degree n if and only if it may be written as a sum of Hermite polynomials of order not more than n:

$$p(x) = \sum_{m=0}^{n} a_m H_m(x)\,.$$

Therefore, (**??**) is true for a given n and all $m < n$ if and only if

$$\int H_n(x) x^m e^{-x^2/2} dx \;=\; 0$$

for all $m < n$. This is the same as saying that $E[H_n(X)X^m] = 0$ for all $m < n$, where X is a standard normal random variable. We compute the first few that way. Of course, $H_0(x) = 1$. $H_1(x)$ is the polynomial of degree 1 that is orthogonal to $X^0 = 1$, which is to say $H_1(x) = x$. Next, $H_2(x)$ is the quadratic that is orthogonal to x and 1. We achieve $E[H_2(X)X] = 0$ if H has the form $x^2 - a$. We achieve $E[H_2(X)X^0] = E[H_2(X)] = 0$ if $a = E[X^2] = 1$, i.e. if $H_2(x) = x^2 - 1$. Finally, we want $H_3(x) = x^3 - ax^2 - bx - c$ to make H_3 orthogonal to 1, x, and x^2. The even powers are automatic if $H_3(x)$ is odd, i.e. if $H_3(x) = x^3 - bx$. The remaining orthogonality condition $E[H_3(X)X] = 0$ is $E[X^4] - bE[X^2] = 0$, which gives $b = 3$, as claimed above. It is the orthogonality properties of Hermite polynomials that makes them handy, for

example, in finding kernel functions for density estimation that satisfy several moment conditions.

In the slightly more general situation of $e^{-x^2/2\rho}$, we can define $H_n(x,\rho) = e^{x^2/2\rho}\partial_x^n e^{-x^2/2\rho}$. These polynomials are orthogonal in the natural sense

$$\int_{-\infty}^{\infty} H_n(x,\rho)H_m(x,\rho)e^{-x^2/2\rho}dx = 0$$

for $m \neq n$. The proof is the same as above.

3.2 Adjoint eigenfunctions

The operator L^* has an *adjoint* operator, L. The eigenvalues of L^* and L are the same but the eigenfunctions are different. The eigenfunctions of L are used to determine the expansion coefficients a_k in (**??**). Later we will find a way to derive the adjoint eigenfunctions directly from the u_n using the principle of *detailed balance*.

We explain the ideas in a finite dimensional setting. Suppose A is an $n \times n$ matrix. The right eigenvectors of A are v_k with $Av_k = \lambda_k v_k$. The left eigenvectors are row vectors u_k with $u_k A = \lambda_k u_k$. This is the same as $A^* u_k^* = \lambda_k u_k^*$. That is, the left eigenvectors of A (when written as column vectors) are the right eigenvectors of the adjoint (transpose) of A. The eigenvalues λ_k are the same. The right and left eigenvectors satisfy *biorthogonality relations* $u_k v_l = 0$ if $\lambda_k \neq \lambda_l$.

For column vectors f and g, let (f,g) be the l^2 inner product $(f,g) = f^*g = \sum_{k=1}^n f_k g_k$. The adjoint of the $n \times n$ matrix A is defined by the property that it should satisfy $(A^*f, g) = (f, Ag)$ for all vectors f and g (check this). It is possible to define other inner products. The one that will be useful for us is the *weighted* l^2 inner product $\langle f,g\rangle_w = \sum_{k=1}^n f_k g_k w_k$, where the w_k are n positive *weights*. Each weighted norm determines a different A^* through the property $\langle A^*f, g\rangle_w = \langle f, Ag\rangle_w$ for all f and g. In this case, the elements of A^* are given by $A^*_{jk} = A_{kj}w_k/w_j$ (check this). The eigenvalues of A^* are the same as the eigenvalues of A (for example, because A^* is similar to the transpose of A). The eigenvectors satisfy the natural biorthogonality relation: If $Av_j = \lambda_j v_j$ and $A^*u_k = \lambda_k u_k$ with $\lambda_j \neq \lambda_k$, then $\langle u_k, v_j\rangle_w = 0$ (check this).

Back to L^*, we compute the adjoint in the L^2 inner product $(f,g) = \int_{-\infty}^{\infty} f(x)g(x)dx$. This adjoint is defined by the requirement that for all functions for which the computation makes sense: $(L^*f, g) = (f, Lg)$. This is done by integration by parts:

$$\begin{aligned}
(L^*f, g) &= \int_{-\infty}^{\infty} \left(L^*f(x)\right) g(x)dx \\
&= \int_{-\infty}^{\infty} \left(\frac{\sigma^2}{2}\partial_x^2 f(x) + \mu\partial_x\left(xf(x)\right)\right) g(x)dx \\
&= \int_{-\infty}^{\infty} f(x)\left(\frac{\sigma^2}{2}\partial_x^2 g(x) - \mu x\partial_x g(x)\right) dx
\end{aligned}$$

$$= \int_{-\infty}^{\infty} f(x)\,(Lg(x))\,dx\ ,$$

with

$$Lg(x) = \frac{\sigma^2}{2}\partial_x^2 g(x) - \mu x \partial_x g(x)\ .$$

Note that the second derivative terms in L^* and L are the same while the first derivative terms change sign. Also, the x is inside the differentiation in L^* but outside in L.

It turns out that the eigenfunctions of L are Hermite polynomials $v_k(x) = H_k(x,\rho) = e^{x^2/2\rho}\partial_x^k e^{-x^2/2\rho}$. The biorthogonality relation $(u_j, v_k) = 0$ for $j \neq k$ is the orthogonality relation for Hermite polynomials, given that $u_j = \partial_x^j e^{-x^2/2\rho} = H_j(x,\rho)e^{-x^2/2\rho}$. The eigenvalue relation $Lv_k = -k\mu v_k$ can be verified by induction on k using the *recurrence relation* (one of many) $v_{k+1}(x) = \frac{-x}{\rho}v_x(x) + \partial_x v_k(x)$.