Simulation of Gaussian Random Variables

Homework 1 due Friday, February 27 at 2 PM.
- explain your work!
- include codes/worksheets for any computational work

Simulation of Gaussian Random Variables

Feel free to use built-in software to generate Gaussian random variables. But you can also DIY with one of two methods:

Box Muller Method

Based on the following property of Gaussian random variables:

If $Y_1, Y_2 \sim N(0,1)$ and independent, then if we change to polar coordinates:

$$Y_1 = R \cos \Theta$$
$$Y_2 = R \sin \Theta$$

Then the random variables $R, \Theta$ are independent, and:

$R^2 \sim \text{Exp}(2), \Theta \sim U(0,2\pi)$. These are easy to simulate, and then one can convert back to Cartesian coordinates. Deriving these properties of $R, \Theta$ is just converting the PDF for $(Y_1, Y_2)$ to $(R, \Theta)$, which is discussed in probability classes and texts.

$$p_{R, \Theta}(r, \theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}}$$

So the implementation of the Box-Muller method is:

1) Simulate $U_1, U_2 \sim U(0,1)$ (independent)
2) Compute $R = \sqrt{-2 \ln U_1}, \Theta = 2\pi U_2$
3) Compute $Y_1 = R \cos \Theta, Y_2 = R \sin \Theta$
This will simulate independent $Y_1, Y_2 \sim N(0,1)$.

To simulate $Z \sim N(\mu, \sigma^2)$, note that $Z \sim \mu + \sigma Y$ where $Y \sim N(0,1)$.

One way to understand this is that deterministic linear operations on (jointly) Gaussian random variables produce Gaussian random variables. That is, the class of Gaussian random variables is a vector space, closed under addition and multiplication by (deterministic) scalar constants.

So we know that $\mu + \sigma Y$ must be Gaussian, so just check its mean and variance.

$$
\mathbb{E}(\mu + \sigma Y) = \mu + \sigma \mathbb{E} Y
$$

$$
\mathbb{E} Y = 0 = \mu
$$

$$
\text{Var}(\mu + \sigma Y) = \text{Var}(\sigma Y) = \sigma^2 \text{Var}(Y)
$$

Variance is invariant under addition by deterministic constants.

More generally, suppose we want to simulate a Gaussian random vector $\bar{Y} \sim N(\bar{\mu}, C)$, where $C$ is the covariance matrix of $\bar{Y}$.

If $C$ is diagonal, then every component of $Y$ is independent and can just be simulated as an independent scalar Gaussian random variable.

Otherwise, if the components of $\bar{Y}$ have nontrivial correlations, then what one should do is find a matrix $S$ such that $S S^T = C$. (There are many possible matrices $S$; any choice works.) This can always be done because $C$ should be a positive definite symmetric matrix, and the standard numerical algorithm for doing this is Cholesky algorithm.

Then you can simulate $Y = \mu + S \bar{\xi}$ where $\bar{\xi} \sim N(0, I)$ (with the same number of components).

Why does this work?
There is one alternative to the Box-Muller method that is sometimes used to simulate Gaussian random variables. The argument for it is that the trigonometric evaluations in the Box-Muller method are sometimes complained to be "expensive." In practice, it's not clear the alternative is faster...but the idea behind it is worth seeing.

**Polar Marsagla method:**

The idea is to avoid the trig evaluations by using the rejection method to generate the appropriate distributions for the polar variables. Here the rejection method, in its primitive form, is used to simulate a uniformly distributed random point in an inconvenient domain by rejection from simulation of a uniformly distributed random point in a surrounding rectangle or rectangular prism.
The definition of a uniformly distributed random variable on a domain $D$ is that for any $A \subset D$, $P(A) = \frac{m(A)}{m(D)}$ where $m(B)$ is the length/area/volume of $B$.

(PDF is uniform constant in Cartesian coordinates)

**Rejection method:** Surround the domain $D$ with a rectangle $R$. Simulate random points in $R$ by simply simulating uniform random numbers along each side of $R$. Accept if in $D$, otherwise reject and try again.

The idea of Polar-Marsagla is that if you can simulate a uniform random point in the unit disc $D$, then the angle coordinate $\theta \sim U(0,2\pi)$ and the radial coordinate $\rho$ can be mapped simply to give the appropriate distribution for the magnitude $R$.

**Implementation of Polar-Marsagla Method:**
1) Use rejection method on the circumscribing square to generate a uniform random point $(V_1, V_2)$ in the unit disc $D$.
2) Map the radius to give the appropriate probability distribution:
   \[ \rho^2 = V_1^2 + V_2^2, \quad h = \sqrt{\frac{-2 \ln \rho^2}{\rho^2}} \]
3) Simulate the random variables as $Y_1 = hV_1, Y_2 = hV_2$. 

\[ \]
Back to the Brownian Motion model, with a view to taking a continuum limit.

To prepare, we use the scaling relationship for Gaussian random variables to write:

\[ \tilde{X}^{(n)} = \sqrt{c \Delta t} \tilde{\xi}^{(n)} \]

where the \( \{ \tilde{\xi}^{(n)} \}^{\infty}_{n=1} \) are iid, \( \sim N(0, I) \). Recall this transformation will give:

\[ \tilde{Z}^{(n)} \sim N(0, c \Delta t I) \]

Then we can write our Brownian motion model

\[ X^{(n+1)}(\omega) = X^{(n)}(\omega) + Z^{(n)}(\omega) \]

as:

\[ \tilde{X}^{(n+1)}(\omega) = \tilde{X}^{(n)}(\omega) + \sqrt{c \Delta t} \tilde{\xi}^{(n)}(\omega) \]

Now we want to contemplate the \( \Delta t \downarrow 0 \) limit, so that the position varies continuously in time. This raises the question of whether we can do that because we said the model was only physically meaningful for \( \Delta t \) larger than the momentum relaxation time scale (\( \sim 10^{-6} \text{ s} \)). The meaning of studying a \( \Delta t \downarrow 0 \) limit for this model, if it is to have any physical validity really means the following:

**momentum relaxation time scale \( \ll \Delta t \ll \text{macroscopic time scale of interest} \)**

\[ 10^{-6} \text{ s} \]

Standard way to take a continuum limit of a discrete-time dynamical model is to try and re-express it in terms of discrete derivatives, then pass to the limit so these become continuous derivatives.
Cannot write down a random version of a standard differential equation to describe the continuum limit of this Brownian motion model. The time rate of change of the state variable does not have a limit, meaning it does not have an instantaneous speed. This is related to the difficulties in measuring the speed of a Brownian particle in the 19th century.

The continuum limit of this simplest Brownian motion model, as well as many other systems with fluctuations "driven by white noise" require the concept of a stochastic differential equation.

The stochastic differential equation describing the continuum limit of our simplest Brownian motion model is:

\[ d\tilde{X}(t, \omega) = \sqrt{c} \, dW(t, \omega) \]

where \( dW(t, \omega) \) is a standardized random process that is just a continuous white noise, meaning that it plays the role of a stream of independent standardized Gaussian random variables, but now indexed by continuous time rather than discrete time (like the discrete-time white noise \( \tilde{\xi}^{(m)}(\omega) \) was). Making sense out of this last sentence requires some care, but that's the intuition.

The way to think about this stochastic differential equation in practice (and for numerical computation) is like writing a differential equation:

\[ \frac{d\tilde{Y}(t)}{dt} = f(\tilde{Y}(t), t) \]

as:

\[ d\tilde{Y}(t) = f(\tilde{Y}(t), t) \, dt \]

which discretizes to:
\[ Y(t + \Delta t) - Y(t) = f(Y(t), t) \Delta t \]

In the same way, the SDE

\[ dX(t, \omega) = \sqrt{\epsilon} dW(t, \omega) \]

is just a continuum limit of the incrementing scheme:

\[ \tilde{X}(t + \Delta t, \omega) - \tilde{X}(t, \omega) = \sqrt{\epsilon} (\tilde{W}(t + \Delta t, \omega) - \tilde{W}(t, \omega)) \]

where \( W(t, \omega) \) is a standard random process known as the Wiener process (or mathematical Brownian motion) which we'll discuss next time.