Math 106 – Notes
Week 8: February 24, 2026

Ethan Levien

Monte Carlo simulation and variance reduction

The goal of Monte Carlo simulation is to approximate expectations with simulations. Consider the expectation

$$\mu(f) = \mathbb{E}_\mu[f(X)] \tag{1}$$

where $f(X) \in L^2(\mu)$. We want to find an estimator $\hat{f}$ of $\mu(f)$ which has a small mean-squared error

$$\text{MSE}_{\hat{f}} = \mathbb{E}[(\hat{f} - \mu(f))^2] \tag{2}$$

This breaks down into Bias and variance:

$$\text{MSE}_{\hat{f}} = \text{var}(\hat{f}) + \left(\mathbb{E}[\hat{f}] - \mu(f)\right)^2 \tag{3}$$

A consistent estimator is one for which $\text{MSE}_{\hat{f}} \to 0$ as $m \to \infty$.

In the simplest case we have iid samples $X_1, \ldots, X_m \sim \mu$ which are used to obtain the approximation

$$\hat{f}_m \approx \frac{1}{m}\sum_{i=1}^m f(X_i) \tag{4}$$

This is an unbiased estimator whose variance is $\text{var}(\hat{f}_m) = \text{var}_\mu(f(X))/m$. Thus,

$$\text{MSE}_{\hat{f}_m} = \text{var}(\hat{f}_m) < \epsilon^2 \implies m > \frac{\text{var}_\mu(f(X))}{\epsilon^2} \tag{5}$$

This idea of variance reduction is to find a new variable $\tilde{X}$ which has the same expectation but a smaller variance.

Importance sampling (Section 4.3)

Importance sampling is one of the main variance reduction techniques. It essentially amounts to performing a change of variables in the expectation in order to sample from a lower variance distribution. Let us suppose that μ has a density and (abusing notation) write $\mu(x)$ for this density. Now let $q(x)$ be another density such that $\mu/q > 0$ for all x. We can write

$$\mathbb{E}_\mu[f(X)] = \mathbb{E}_q\left[f(X)L(X)\right], \quad L(X) = \frac{\mu(X)}{q(X)} \tag{6}$$

This suggests the estimator

$$\hat{f}_{m,q} \approx \frac{1}{m}\sum_{i=1}^m f(\tilde{X}_i)L(\tilde{X}_i) \tag{7}$$

where $\tilde{X}_i$ are iid samples from q. This works to our advantage if

$$\text{var}_q(f(X)L(X)) < \text{var}_\mu(f(X)). \tag{8}$$

Motiving Example: Estimating the time to a rare event

Now let's see how we can perform a change of measure to perform importance sampling on the path space with a concrete example. Consider again an OU process,

$$dX_t = -\theta X_t \, dt + \sqrt{2D} \, dW_t, \tag{9}$$

and suppose we are interested in an event that occurs with (state-dependent) rate $\lambda(X_t)$. We assume λ is an increasing function of x. Conditional on $\mathcal{F}_t^X$, the survival probability is

$$\mathbb{P}(\tau > t \mid \mathcal{F}_t^X) = \exp\left(-\int_0^t \lambda(X_s) \, ds \right), \tag{10}$$

and therefore

$$\mathbb{P}(\tau > t) = \mathbb{E}_\mu \left[\exp\left(-\int_0^t \lambda(X_s) \, ds \right) \right], \tag{11}$$

where μ denotes the law of the path $\{X_s\}_{0 \le s \le t}$.

We would like to sample from another SDE where τ happens sooner (and hopefully we don't pay a price in the simulation cost).

Detour: Radon–Nikodym derivatives

To make use of importance sampling in the context of SDEs we need to understand when a change of measure on *path space* is valid. In finite dimensions we often write densities $\mu(x)$ and $q(x)$, but for diffusion processes the relevant objects are probability measures on a space of paths. In this setting, the analogue of having a density is *absolute continuity*.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t))$ be a filtered measurable space, and let μ and q be probability measures on $(\Omega, \mathcal{F})$.

Definition 1. *We say that μ is absolutely continuous with respect to q if*

$$\mu \ll q \tag{12}$$

if $q(A) = 0$ implies $\mu(A) = 0$ for all $A \in \mathcal{F}$.

In this case, the Radon–Nikodym theorem guarantees the existence of a nonnegative observable L such that

$$\mu(A) = \int_A L(x) q(dx) \qquad \text{for all } A \in \mathcal{F}. \tag{13}$$

We then write $L = d\mu/dq$ and call it the Radon–Nikodym derivative (or likelihood ratio). Equivalently, for a random variable f,

$$\mu(f) = \mathbb{E}_\mu[f] = \mathbb{E}_q[f\, L] = q(fL). \tag{14}$$

In applications to SDEs we typically consider $\mathcal{F}_t$ (the information up to time t) and require absolute continuity *locally in time*,

$$\mu\big|_{\mathcal{F}_t} \ll q\big|_{\mathcal{F}_t}, \tag{15}$$

so that a likelihood ratio $L_t = d\mu/dq|_{\mathcal{F}_t}$ exists as an adapted process.

What goes wrong when changing measures on path space

As an example, consider two OU processes on $[0, T]$:

$$dX_t = -\theta X_t\, dt + \sqrt{2D}\, dW_t^{(1)}, \tag{16}$$

$$dY_t = -\theta Y_t\, dt + \sqrt{2\tilde{D}}\, dW_t^{(2)}, \tag{17}$$

where $W^{(1)}$ and $W^{(2)}$ are Wiener processes (possibly on different probability spaces). Let μ and q denote the laws on the path space of X_t and Y_t respectively.

A quick way to see what goes wrong is to consider the case $D \neq \tilde{D}$. As I showed in class, the quadratic variations of X_t and Y_t are

$$[X, X]_T = 2DT \qquad [Y, Y]_T = 2\tilde{D}T \tag{18}$$

respectively where the equalities should be interpreted in the almost sure sense (meaning they occur with probability 1 under μ and q respectively). Define the measurable event in path space

$$A := \{\omega \in C([0, T]) : [\omega]_T = 2DT\}. \tag{19}$$

Here ω could be replaced by X or Y but we write it this way to indicate this is an event that can be considered under either μ or q. Then

$$\mu(A) = 1, \qquad q(A) = 0 \quad \text{if } \tilde{D} \neq D, \tag{20}$$

Therefore μ is *not* absolutely continuous with respect to q. In fact the measures are mutually singular, meaning there exists an event of probability one under μ and probability zero under q (and vice versa). Consequently, the Radon–Nikodym derivative $d\mu/dq$ does not exist on path space when $D \neq \tilde{D}$, and there is no likelihood ratio that can convert μ–expectations into q–expectations for general path functionals.

Note that even when $D \neq \tilde{D}$, the finite-dimensional distributions at fixed times $\{t_1, \ldots, t_n\}$ are mutually absolutely continuous (both are nondegenerate Gaussians on $\mathbb{R}^n$). The problem, however, is that the likelihood ratio blows up as $dt \to 0$. To see this, consider the finite-dimensional distribution of $\sqrt{2D}W_t$ at times $0, dt, 2dt, \ldots, ndt$:

$$\rho_D(w_1, \ldots, w_n) = \prod_{i=1}^{n} (4\pi D dt)^{-1/2} e^{-(\Delta w_i)^2/(4Ddt)} = (4\pi D dt)^{-n/2} e^{-1/(4Ddt)\sum_{i=1}^{n}(\Delta w_i)^2} \tag{21}$$

The ratio

$$\frac{\rho_D(w_1, \ldots, w_n)}{\rho_{\tilde{D}}(w_1, \ldots, w_n)} = \left(\tilde{D}/D\right)^{n/2} e^{(\tilde{D}^{-1} - D^{-1})\sum_{i=1}^{n}(\Delta w_i)^2/(4dt)}. \tag{22}$$

Although dt has canceled in the ratio, there is still a factor of $1/dt$ in the exponent which causes it to blow-up as $dt \to 0$.

Changing the drift and Girsanov's Theorem

Since we cannot change the diffusion coefficient, we need to adjust the drift term. Returning to the context of our modeling example – that is, the problem of estimating the CDF of τ – let's explore this possibility. To this end, let $\{u_t\}_{0 \leq t \leq T}$ be an adapted control (often $u_t = u(X_t)$). Define a new probability measure q on $\mathcal{F}_t$ under which X solves

$$dX_t = \Big(-\theta X_t + \sqrt{2D}\, u_t \Big)\, dt + \sqrt{2D}\, dW_t^{(q)}. \tag{23}$$

Let q denote the induced law of the path $\{X_s\}_{0\le s\le t}$.

Using the Euler–Maruyama approximation (we will eventually take the limit $dt \to 0$, so the order of the approximation doesn't matter), we get

$$p_q(x_i \mid x_{i-1}) = \frac{1}{\sqrt{4\pi D\, dt}} \exp\left(-\frac{\left(x_i - x_{i-1} + \theta x_{i-1} dt - \sqrt{2D}\, u_{i-1} dt\right)^2}{4D\, dt} \right), \tag{24}$$

while under the original OU law μ,

$$p_\mu(x_i \mid x_{i-1}) = \frac{1}{\sqrt{4\pi D\, dt}} \exp\left(-\frac{\left(x_i - x_{i-1} + \theta x_{i-1} dt\right)^2}{4D\, dt} \right). \tag{25}$$

Since the prefactors match, their ratio is

$$\frac{\rho_\mu(x_1,\ldots,x_n \mid x_0)}{\rho_q(x_1,\ldots,x_n \mid x_0)} = \exp\left\{ -\frac{1}{4D\, dt} \sum_{i=1}^n \left[\left(\Delta x_i + \theta x_{i-1} dt\right)^2 - \left(\Delta x_i + \theta x_{i-1} dt - \sqrt{2D}\, u_{i-1} dt\right)^2 \right] \right\}, \tag{26}$$

Using $(a)^2 - (a-b)^2 = 2ab - b^2$ with $a = \Delta x_i + \theta x_{i-1} dt$ and $b = \sqrt{2D}\, u_{i-1} dt$, we obtain

$$\log \frac{\rho_\mu}{\rho_q} = -\frac{1}{4D\, dt} \sum_{i=1}^n \left[2\left(\Delta x_i + \theta x_{i-1} dt\right)\left(\sqrt{2D}\, u_{i-1} dt\right) - \left(\sqrt{2D}\, u_{i-1} dt\right)^2 \right] \tag{27}$$

$$= -\sum_{i=1}^n u_{i-1} \frac{\Delta x_i + \theta x_{i-1} dt}{\sqrt{2D}} + \frac{1}{2} \sum_{i=1}^n u_{i-1}^2\, dt. \tag{28}$$

Now, under q the Euler increment can be written as

$$\Delta x_i + \theta x_{i-1} dt = \sqrt{2D}\, u_{i-1} dt + \sqrt{2D}\, \Delta W_i^{(q)} \tag{29}$$

so

$$\log \frac{\rho_\mu}{\rho_q} = -\sum_{i=1}^n u_{i-1} \left(u_{i-1} dt + \Delta W_i^{(q)} \right) + \frac{1}{2} \sum_{i=1}^n u_{i-1}^2\, dt \tag{30}$$

$$= -\sum_{i=1}^n u_{i-1}\, \Delta W_i^{(q)} - \frac{1}{2} \sum_{i=1}^n u_{i-1}^2\, dt. \tag{31}$$

Therefore the finite-dimensional likelihood ratio is

$$\frac{\rho_\mu(x_1,\ldots,x_n \mid x_0)}{\rho_q(x_1,\ldots,x_n \mid x_0)} = \exp\left(-\sum_{i=1}^n u_{i-1}\, \Delta W_i^{(q)} - \frac{1}{2} \sum_{i=1}^n u_{i-1}^2\, dt \right). \tag{32}$$

So the continuous-time Radon–Nikodym derivative is

$$L_t \coloneqq \frac{d\mu}{dq}\Big|_{\mathcal{F}_t} = \exp\left(-\int_0^t u_s\, dW_s^{(q)} - \frac{1}{2} \int_0^t u_s^2\, ds \right). \tag{33}$$

Combining everything:

$$\mu(f) = q(f L_t) = \mathbb{E}_q\left[\exp\left(-\int_0^t \lambda(X_s)\, ds - \int_0^t u_s\, dW_s^{(q)} - \frac{1}{2} \int_0^t u_s^2\, ds \right) \right] \tag{34}$$

This is an application of *Girsanov's theorem*, which tells us how to change measure of Wiener paths.

Precise statement of Girsanov's Theorem 1

Theorem 1 (Special case of T9.2)**.** *Consider the Itô process*

$$d\tilde{W}_t = \phi(t,\omega)dt + dW_t, \quad \tilde{W}_0 = 0 \tag{35}$$

defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$*. Assume*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |\phi(s,\omega)|^2 ds\right)\right] < \infty \tag{36}$$

Now let $\tilde{\mathbb{P}}$ *be another probability measure on* $(\Omega, \mathcal{F})$ *defined by*

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\Big|_{\mathcal{F}_t} = \exp\left(-\int_0^t \phi(s,\omega)dW_s - \frac{1}{2}\int_0^t \phi(s,\omega)^2 ds\right) \tag{37}$$

Then $\tilde{W}_t$ *is a Wiener process under* $\tilde{\mathbb{P}}$*.*

Remark 1. *To connect to our notation, the distribution of* W_t *under* $\mathbb{P}$ *is* μ *–* μ *is the law of a Wiener process.* $\mathbb{P}$ *induces a distribution over* $\tilde{W}_t$ *which is our* q*.* q *is not the law of a Wiener process because of the drfit term, but the change of measure converts it back to one.*

Feynman-Kac formula (Section 8.6)

In the previous example we encountered the path expectation

$$\mathbb{E}_\mu\left[\exp\left(-\int_0^t \lambda(X_s)\, ds\right)\right] \tag{38}$$

The challenge of sampling from these expectations is pervasive. Because this is not the expectation of X_t at a fixed time conditional on the initial point, it does not obey the usual backward equation. It does however have a semigroup structure.

$$v(x,t) = \mathbb{E}_\mu\left[f(X_t)\exp\left(-\int_0^t \lambda(X_s)\, ds\right)|X_0 = x\right]. \tag{39}$$

Note that if we define $Z_t = Z_0 + \int_0^t \lambda(X_s)\, ds$, we have a new system of SDEs

$$dZ_t = \lambda(X_t)dt \tag{40}$$

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{41}$$

with this definition of Z_t. This has the generator

$$(\mathcal{A}g)(x,z) = (\mathcal{L}g)(x,z) + \lambda(x)\partial_z g(x,z) \tag{42}$$

If we consider

$$w(x,z,t) = \mathbb{E}_\mu[f(X_t)\exp(-Z_t) \mid X_0 = x, Z_0 = z] \tag{43}$$

Then, with $g(x,z) = f(x)e^{-z}$, we have

$$\partial_t w(x,z,t) = \mathcal{A}w(x,z,t), \quad w(x,z,0) = f(x)e^{-z} \tag{44}$$

Now observe that $v(x,t) = w(x,0,t)$, hence

$$\partial_t v(x,t) = \mathcal{L}v(x,t) + \lambda(x)\partial_z w(x,z,t)\Big|_{z=0} \tag{45}$$

Due to

$$\partial_z w(x,z,t) = \mathbb{E}_\mu\left[f(X_t)\,\partial_z \exp\left(-\int_0^t \lambda(X_s)\,ds - z\right)\;\Big|\; X_0 = x\right] \tag{46}$$

$$= -\mathbb{E}_\mu\left[f(X_t)\,\exp\left(-\int_0^t \lambda(X_s)\,ds - z\right)\;\Big|\; X_0 = x\right]. \tag{47}$$

hence

$$\partial_z w(x,0,t) = -v(x,t). \tag{48}$$

and therefore we arrive at the *Feynman-Kac* PDE

$$\partial_t v(x,t) = \mathcal{L}v(x,t) - \lambda(x)v(x,t). \tag{49}$$

In particular, with $v(x,0) = 1$, this becomes

$$v(x,t) = \mathbb{E}_\mu\left[\exp\left(-\int_0^t \lambda(X_s)ds\right)\Big|X_0 = x\right] \tag{50}$$

The problem of computing $\mu_t(f) = \mathbb{E}[v(X_0,t)]$ has thus transformed into a problem of solving a PDE.

Particle interpretation of Feynman-Kac formula

In the case of Brownian dynamics we gave an interpretation of the density from a many-particle view. In particular, we viewed the probability density as describing a normalized concentration for a large particle system. The same can be done for the Feynman-Kac PDE, but we now need to view it as relating to densities in a system with changing particle number.

To see this, consider a population of particles which die at rate $\lambda(X_t)$. Define $N(X_t \in B \mid X_0)$ to be the number of particles alive whose positions are in a measurable set B and started from initial point X_0. We then define a number density $n(x,t)$, similarly to a probability density, as

$$N(X_t \in B \mid X_0) = \int_B n(x,t)\,dx \tag{51}$$

Consider how $n(x,t)$ changes due to flux and particle death; by a conservation of mass argument similar to the one we used when deriving the flux for Brownian dynamics, we have

$$\partial_t n(x,t) = -\nabla \cdot j(x,t) - \lambda(x)n(x,t) = \mathcal{L}^* n(x,t) - \lambda(x)n(x,t). \tag{52}$$

We can then define

$$v_t(f) = \int n(x,t)\,f(x)\,dx \tag{53}$$

Thus $v_t(f)$ has the adjoint equation. We can create a normalized density associated with n to obtain population expectations. This means that another way to obtain $\mu(f)$ is via a population simulation. This is generally more stable than evaluating the expectation directly.