Summer School in Peking University
Computational Methods for Rare Events

Professor: Jonathan Weare
Scribe: Alex

Contents

1 Introduction

In this lecture, we will talk about rare event simulation. Rare events are usually the most important characteristics of a system to predict and understand. For example, Airplane manufacturing, extreme weather, drugs...

We can try to estimate RE statistics, for example occurrance probabilities, by:

- analysing existing observational data
- use a computer model (npde) to generate synthetic data.

The first method is hard because few examples exist in the history record. The second is also hard because it is too expensive to generate enough data for accurate modeling (short timescale). We will mostly focus on the second method.

We discuss the problem in two parts:

- The Static setting: We are given a probability density π and are asked to estimate:

$$\pi[f] := \mathbb{E}_X(f(X)) = \int f(x)\pi(x)\,\mathrm{d}x.$$

We are particularly interested in cases in regions where f is large and π is small.

The dynamics setting: Same as teh static case but now X is the path of Markov process and f is a functional of the path. In this setting we may not have direct access to π. We may not even have the rules that govern the evolution of X (i.e. X be a "black box").

Some basic Monte Carlo concepts: Suppose we can generate independent identically distributed (i.i.d) samples $X^{(j)}$ from π. The simplest estimation of $\pi[f]$ is:

$$\bar{f}_M = \frac{1}{M}\sum_{j=1}^{M} f(X^{(j)}).$$

$\bar{f}_M$ is an unbiased estimation. We hope that $\bar{f}_M \to \pi[f]$ as $M \to \infty$. At least we want $\bar{f}_M \xrightarrow{P} \pi[f]$. We often expect that mean square convergence: $\mathbb{E}[|\bar{f}_M - \pi[f]|^2] \xrightarrow{M\to\infty} 0$.

1.1 Monte Carlo Estimation

We can decompose the MSE into variance term and bias term:

$$\mathbb{E}[|Y - \pi[f]|^2] = \underbrace{\mathbb{E}[|Y - \mathbb{E}(Y)|^2]}_{var} + \underbrace{\mathbb{E}[|\mathbb{E}(Y) - \pi(f)|^2]}_{bias}.$$

and $\mathbb{E}[|Y - \mathbb{E}(Y)|^2] \le \frac{\mathrm{Var}(f)}{M}$

We can also often prove a concentration bound (Chebychev). Define the moment equality function: $\Lambda(t) = \log \mathbb{E}_\pi[e^{tf(X)}]$. By Markov inequality:

$$\mathbb{P}(\bar{f}_M - \pi[f] > \epsilon) = \mathbb{P}(e^{(\bar{f}_M - \pi[f])Mt} > e^{Mt\epsilon}) \le \frac{\mathbb{E}[e^{t(f(X)-\pi[f])}]^M}{e^{Mt\epsilon}} = e^{M(t(\epsilon+\pi[f])-\Lambda(t))} = e^{-M\gamma(\epsilon)}$$

where $\gamma(\epsilon) := \sup_{t>0}\{t(\epsilon + \pi[f]) - \Lambda(t)\}$.

Since $\pi[f]$ is small, we want to control the relative accuracy, e.g. $\mathbb{P}\left(\frac{\bar{f}_M - \pi[f]}{\pi[f]}\right)$ or $\log \bar{f}_M - \log \pi[f]$. Notice that:

$$\frac{\mathrm{Var}(\bar{f}_M)}{\pi[f]^2} = \frac{1}{M}\left(\frac{\pi[f^2]}{\pi[f]^2} - 1\right),$$

and if $f = 1_B(x)$, then:

$$\frac{\mathrm{Var}(\bar{f}_M)}{\pi[f]^2} = \frac{1}{M}\left(\frac{1}{\pi[f]} - 1\right).$$

I must remark here that our target is to reduce variance.

1.2 An asymptotic "rare event" setting

Let V, g be upper semicontinuous functions. Assume they are bounded below and that:

$$\liminf_{\|X\|\to\infty} \frac{f(X)}{\|X\|} = \infty \quad \text{and} \quad \inf_X V(X) = 0$$

2 Importance Sampling

The higher importance, the higher probability we sample it. Say, suppose $Y^{(1)}, ..., Y^{(M)}$ are independent and drawn from a reference density $\tilde{\pi}$:

$$\tilde{f}_M = \frac{1}{M}\sum_{j=1}^{M} f(Y^{(j)})\frac{\pi(Y_j)}{\tilde{\pi}(Y_j)}$$

What we care about is:

$$\mathrm{Var}(\tilde{f}_M) = \frac{1}{M}\mathrm{Var}_{\tilde{\pi}}(f\frac{\pi}{\tilde{\pi}})$$

We see, in the region where π is large $\tilde{\pi}$ is small while f is small too. What π is the best is the question.

Of course we want to reduce variance:

$$\mathrm{Var}(\bar{f}_M) = \frac{1}{M}\left(\int f(y)^2\frac{\pi(y)^2}{\tilde{\pi}(y)^2}\tilde{\pi}(y)dy - \pi[f]^2\right).$$

With the Cauchy inequality, we have:

$$\int f(y)^2\frac{\pi(y)^2}{\tilde{\pi}(y)^2}\tilde{\pi}(y)\,\mathrm{d}y \geq (\int f(y)\pi(y)dy)^2 = \pi[f]^2,$$

the condition for equivalence is $\tilde{\pi}(x) = \frac{|f|(x)\pi(x)}{\pi[f]}$.

2.1 Autonormalized Importance Sampling

Notice $1_M = \frac{1}{M}\sum_{j=1}^M f(Y^{(j)})\frac{\pi(Y^{(j)})}{\tilde{\pi}((Y^{(j)})}$, and then we have:

$$\frac{\tilde{f}_M}{\tilde{1}_M} = \frac{\sum_{j=1}^M f(Y^{(j)})\frac{\pi(Y^{(j)})}{\tilde{\pi}((Y^{(j)})}}{\sum_{j=1}^M \frac{\pi(Y^{(j)})}{\tilde{\pi}((Y^{(j)})}}.$$

2.2 Laplace Principal

Let $\pi_\epsilon(x) \sim e^{-\frac{V(x)}{\epsilon}}$ and $f_\epsilon(x) = e^{-\frac{g(x)}{\epsilon}}$ for $0 < \epsilon \ll 1$.
For example: $V(x) = \frac{\|x\|^2}{2}$, $\pi_\epsilon(x) = \mathcal{N}(0, \sqrt{\epsilon})$.
For example: If $B \subset \mathbb{R}^d$, $g(x) = \begin{cases} 0, & \text{if } x \in B, \\ \infty, & \text{if } x \notin B \end{cases}$, then $f_\epsilon(x) = 1_B(x)$.
For example: If f_ϵ is a data likelihood and π_ϵ is a prior, then:

$$p(x|\mathrm{data}) = \frac{p(\mathrm{data}|x)p(x)}{p(\mathrm{data})} = \frac{f_\epsilon(x)\pi_\epsilon(x)}{\int f_\epsilon(x)\pi_\epsilon(x)dx}$$

We will see that $\pi_\epsilon[f_\epsilon]$ is usually hard to estimate for small ϵ. (Maybe here we can view ϵ as the rareness of the events.)

Laplace Principal: For any semi continuous function h with $\lim_{\|X\|\to\infty}\frac{h(X)}{\|X\|} = \infty$, and $h_* = \inf h(x) > -\infty$, then:

$$\lim_{\epsilon\to 0}\epsilon\log\int e^{-h(x)/\epsilon}\,\mathrm{d}x = -h_*.$$

2.3 Saddle point approximation

Assume h is C^2 and has a unique global minimum at x_h. Suppose the second derivative matrix $D^2h(x_h)$ is strictly positive definite. Then:

$$\int e^{-h(x)/\epsilon}\,\mathrm{d}x = \epsilon^{d/2}e^{-h(x_h)/\epsilon}\left(\frac{(2\pi)^{d/2}}{|D^2h(x_h)|^{1/2}} + O(1)\right)$$

Proof: We calculate the $\int_{B_r} e^{-(h(x)-h_*)/\epsilon}\,\mathrm{d}x$ using Taylor's expansion!

We can use the Laplace Principal to estimate the rate of decay of $\pi_\epsilon[f_\epsilon]$. If V and g are upper-semi continuous and bounded below. Then:

$$\epsilon \log e^{-V(x)/\epsilon} e^{-g(x)/\epsilon} \to -\inf_x (V+g),$$

and observe that:

$$\epsilon \log \int e^{-V(X)/\epsilon} \, \mathrm{d}X \to \inf_x V(X) = 0,$$

then:

$$\epsilon \log \pi_\epsilon[f_\epsilon] = \epsilon \log e^{-V(x)/\epsilon} e^{-g(x)/\epsilon} - \epsilon \log \int e^{-V(X)/\epsilon} \, \mathrm{d}X \to -\inf_x (V+g)$$

Recall that $\frac{\mathrm{Var}(\bar{f}_M)}{\pi[f]^2} = \frac{1}{M}(\frac{\pi[f^2]}{\pi[f]^2} - 1)$, by the analysis above: we have:

$$\epsilon \log \left(\frac{\mathrm{Var}(\bar{f}_M)}{\pi[f]^2} \right) \to \inf_x (2V + 2g) - \inf_x (V + 2g) =: \gamma \geq 0.$$

So the relative MSE of the $\bar{f}_M$ increase exponenially with $1/\epsilon$.

Recall that the optimal IS estimator uses $\tilde{\pi} = \frac{e^{-g/\epsilon} e^{-V/\epsilon}}{\pi_\epsilon[f_\epsilon]}$, and $\tilde{\pi}[f_\epsilon^2 (\frac{\pi_\epsilon}{\pi})^2] = \pi_\epsilon^2 \to 0$ as $\epsilon \to 0$.

We cannot use $\tilde{\pi}$ however we can find a more practical $\tilde{\pi}$, which means:

$$\epsilon \log \left(\frac{\mathrm{Var}(\bar{f}_M)}{\pi[f]^2} \right) \to 0.$$

From now, we assume $V \in C^1$.

Here we introduce the exponential twist, which means:

$$\tilde{V}(x) = V(x) - w^T x - \inf_x \{V(x) - w^T x\}$$

for some $w \in \mathbb{R}^d$. So $\tilde{\pi}_\epsilon(x) = \frac{\pi_\epsilon(x) e^{w^T x}}{\pi_\epsilon[e^{w^T x}]}$. In fact, $\tilde{\pi}_\epsilon$ is a mean shift of π_ϵ. Similarly, $\epsilon \log \left(\frac{\mathrm{Var}(\bar{f}_M)}{\pi[f]^2} \right) \to \inf_x \{2g + 2V\} - \inf_x \{2g + 2V - \tilde{V}\} = \tilde{\gamma}(w) \geq 0$. We want to choose w so that $\tilde{\gamma}(w) = 0$.

Let x_{g+V} be a global minimum of $g + V$ and we will choose $w = \nabla V(x_{g+V})$. Notice that $\nabla \tilde{V}(x_{g+V}) = \nabla V(x_{g+V}) - w = 0$. That means x_{g+V} is also a minimum of $\tilde{V}$. If we assume that:

- x_{g+V} is the global minimum of $\tilde{V}$, so $\tilde{V}(x_{g+V}) = 0$, this holds, for example, V is convex.
- x_{g+V} is also a global minimum of $2g + 2V - \tilde{V}$ so that $\inf_x \{2g + 2V - \tilde{V}\} = 2g(x_{g+V}) + 2V(x_{g+V}) - \tilde{V}(x_{g+V}) = 2g(x_{g+V}) + 2V(x_{g+V})$

and the assumptions presented above ensure $\tilde{\gamma} = 0$.

Recall that:

$$\frac{1}{M} \log \mathbb{P}(\bar{f}_M - \pi[f] > x) = -\underbrace{\sup_{t>0} \{t(x + \pi[f]) - \log \mathbb{E}_\pi[e^{tx}]\}}_{I(x)}$$

let $V \to I(x)$ and $\epsilon \to \frac{1}{M}$. (The derivation requires the knowledge of large deviation theory, which leads to similar results.)

An important aspect of the work is to explore the presence of multiple minima, as this will lead to changes in the regimes:

$$\tilde{\pi} = \sum_{i=1}^m q_i \tilde{\pi}_i,$$

where $\sum_{i=1}^m q_i = 1$.

3 Stratification

The motivation here is that we can choose more samples which are more significant to us instead of just sampling from the distribution.

Last time we mentioned of the form:

$$\tilde{\pi} = \sum_{i=1}^{m} q_i \tilde{\pi}_i,$$

and we further define M_i as the number of samples from $\tilde{\pi}_i$, we have:

$$f_M = \sum_{i=1}^{m} \frac{M_i}{M} \frac{1}{M_i} \sum_{j=1}^{M_i} \frac{f(Y_i^{(j)})}{\sum_{l=1}^{m} q_l \frac{\tilde{\pi}_l(Y_i^{(j)})}{\pi(Y_i^{(j)})}},$$

and if M_i are chosen in advance, then the related estimator is:

$$f_M^S = \sum_{i=1}^{m} q_i \frac{1}{M_i} \sum_{j=1}^{M_i} \frac{f(Y_i^{(j)})}{\sum_{l=1}^{m} q_l \frac{\tilde{\pi}_l(Y_i^{(j)})}{\pi(Y_i^{(j)})}},$$

Remark: $(q_1, ..., q_l) = (1, 0, ..., 0)$, and we have $f_M^S = f_M$.

Of course the estimator is unbiased, i.e.

$$\mathbb{E}[f_M^S] = \pi[f].$$

And the variance:

$$\mathrm{Var}[f_M^S] = \sum_{i=1}^{m} \frac{q_i^2}{M_i} \left(\tilde{\pi}_i[f^2(\frac{\pi}{\tilde{\pi}})^2] - (\tilde{\pi}_i[f(\frac{\pi}{\tilde{\pi}})])^2 \right).$$

Assume that $M_i \approx M q_i$:

$$\sum_{i=1}^{m} \frac{q_i^2}{M_i} \left(\tilde{\pi}_i[f^2(\frac{\pi}{\tilde{\pi}})^2] \right) \approx \frac{1}{M} \sum_{i=1}^{m} q_i \left(\tilde{\pi}_i[f^2(\frac{\pi}{\tilde{\pi}})^2] \right) = \frac{1}{M} \sum_{i=1}^{m} \tilde{\pi}[f^2(\frac{\pi}{\tilde{\pi}})^2],$$

and:

$$\sum_{i=1}^{m} \frac{q_i^2}{M_i} \left((\tilde{\pi}_i[f(\frac{\pi}{\tilde{\pi}})])^2 \right) \approx \frac{1}{M} \sum_{i=1}^{m} q_i (\tilde{\pi}_i[f(\frac{\pi}{\tilde{\pi}})])^2 \overset{\text{AM-inequality}}{\geq} \frac{1}{M} \left(\sum_{i=1}^{m} q_i \tilde{\pi}_i[f(\frac{\pi}{\tilde{\pi}})] \right)^2 = \frac{1}{M} (\pi[f])^2.$$

So we have:

$$\mathrm{Var}[f_M^S] \leq \mathrm{Var}[f_M].$$

To use either of these estimators we need to be able to evaluate $\frac{\tilde{\pi}_l}{\pi}$. We write: $\tilde{\pi}_l \propto \tilde{p}_l$ and $\pi \propto p$, and $c_l = \frac{\int \tilde{p}_l \, \mathrm{d}x}{\int p \, \mathrm{d}x}$. And we let $\bar{q}_i = \frac{q_i}{c_i}$ which is known:

$$f_M^S = \sum_{i=1}^{m} c_i \bar{q}_i \frac{1}{M_i} \sum_{j=1}^{M_i} \frac{f(Y_i^{(j)})}{\sum_{l=1}^{m} \bar{q}_l \frac{\tilde{p}_l(Y_i^{(j)})}{p(Y_i^{(j)})}}.$$

Recall that $\pi[\frac{\tilde{p}_l}{p}] = c_l$, we can calculate:

$$c_j \bar{q}_j = \bar{q}_j \pi[\frac{\tilde{p}_j}{p}] = \bar{q}_j \mathbb{E}\left[\left(\frac{\tilde{p}_j}{p} \right)_M^S \right] = \bar{q}_j \sum_{i=1}^{m} c_i \bar{q}_i \pi_i \left[\frac{\tilde{p}_j / p}{\sum_{l=1}^{m} \bar{q}_l \tilde{p}_l / p} \right] = \sum_{i=1}^{m} c_i \bar{q}_i \pi_i \left[\frac{\bar{q}_j \tilde{p}_j}{\sum_{l=1}^{m} \bar{q}_l \tilde{p}_l} \right].$$

We define $u_j = c_j \bar{q}_j$ and $F_{ij} = \pi_i \left[\frac{\bar{q}_j \tilde{p}_j}{\sum_{l=1}^m \bar{q}_l \tilde{p}_l} \right]$, then we'll give:

$$u^T = u^T F.$$

In conclusion, we can use f_S^M as below:

- Generate M_i samples Y_i from $\tilde{p}_i$.
- Solve $u^T = u^T F$.
- Assemble the estimator $f_M^S = \sum_{i=1}^m c_i \bar{q}_i \frac{1}{M_i} \sum_{j=1}^{M_i} \frac{f(Y_i^{(j)})}{\sum_{l=1}^m \bar{q}_l \frac{\tilde{p}_l(Y_i^{(j)})}{p(Y_i^{(j)})}}$.

3.1 The RE Settings

Suppose: $\pi_\epsilon \propto e^{-V/\epsilon}$, $\tilde{\pi}_i \propto e^{V_i/\epsilon}$, $f_\epsilon = e^{-g/\epsilon}$.

To make things simpler, we assume $V(x) = \infty$, for $x \in [-R, R]^d$, and that g is only a function of $x_1 \in [-R, R]$. Informally, the stratified scheme is just trying to achieve $\inf_x\{2g + 2V\} - \inf_x\{2g + 2V - \tilde{V}\} = O(\epsilon)$.

The sketch of the idea is to chop up the distribution into intervals with length of ϵ.

In conclusion, we have the methods below:

- $\bar{f}_M$ doesn't work for rare events.
- $\tilde{f}_M$ with optimal $\tilde{\pi}$ has zero variance but not practical.
- $\tilde{f}_M$ with exponential twists is more practical and work for rare events under the separating hyperplane conditions.
- $\tilde{f}_M$ with a mixture of twists.
- f_M^S is usually better than $\tilde{f}_M$ and leads to an eigenequation of c_l.

Our target is to reduce variance!

References: "Stratification as a general variance reduction strategy" in JUQ.

4 Rare events for Markov Processes

$X_t \in \mathbb{R}^d$ is a sequence of R.Vs indexed by $t \in \mathbb{R}_{\geq 0}$ or $\mathbb{Z}_{\geq 0}$. X_t is a Markov Process, if for any $s < t$:

$$P(X_t | \mathscr{F}_s) = P(X_t | X_s),$$

where $\mathscr{F}_s$ is the fitration generated by $\{X_r\}_{r \leq s}$.

If $X_t \in [n]$, the evolution of X_t is generated by:

$$T_{ij} = P(X_{t+1} = j | X_t = i).$$

For $g \in \mathbb{R}^n$:

$$(Tg)_i = \sum_j P(X_{t+1} = j | X_t = i) g_j = \mathbb{E}[g X_{t+1} | X_t],$$

and

$$(\nu^T T)_j = \sum_i P(X_{t+1} = j | X_t = i) \nu_i = P(X_{t+1} = j),$$

if $X_t \sim \nu$.

The invariant prob-vector ν if:

$$(\nu^T T)_j = \nu_j.$$

And recall the power method,

$$\rho_n = \frac{A\rho_{n-1}}{\|A\rho_{n-1}\|},$$

we have $\lim_{n\to\infty} \rho_n =$ the largest eigenvector of the A.

Let's now talk about a distribution:

$$\nu_j = e^{cos(\frac{2\pi(j-1)}{n})/\epsilon},$$

$T_{i,i\pm1} = \frac{\nu_{i\pm1}}{2(\nu_i+\nu_{i\pm1})}$, and $T_{ii} = 1 - T_{i,i-1} - T_{i,i+1}$. With the decay of the ϵ, the distribution concentrates on 1 and n. The "1 and n" is the normal situations while the $\frac{n-1}{2}$ is the extreme situations. We care about how the extreme situations happen!

When $\nu_{i\pm1} > \nu_i$, the process prefers to move from i to $i\pm1$. We might be interested in estimating the probability that X_t reaches n before it reaches 1. We claim that if X_0 is near 1, this probability is exponential small in $\frac{1}{\epsilon}$ (which is a random event), if we try to estimate this probability using the basic MC estimator:

$$\bar{f}_M = \frac{1}{M}\sum_{j=1}^{M} 1_{\{X_t^{(j)}\text{hits } n \text{ before it hits } 1\}},$$

where $X_t^{(j)}$ are independent realization of X_t. For exactly the same reason, we have:

$$\frac{\text{Var}(\bar{f}_M)}{P\{X_t^{(j)}\text{hits } n \text{ before it hits } 1\}^2} = \frac{1}{M}\left(\frac{1}{P\{X_t^{(j)}\text{hits } n \text{ before it hits } 1\}} - 1\right) = \text{exponential large in } \frac{1}{\epsilon}.$$

T and the distribution of X_0 completely determine the distribution of $\{X_r\}_r$.

Often, T is a kernel operator:

$$Tg(x) = \int g(y)q(y|x)\,\mathrm{d}y,$$

and:

$$\nu(T(B)) = \int_{y\in B}\int_x q(y|x)\nu(x)\,\mathrm{d}x\,\mathrm{d}y.$$

For example, the diffusion can be defined as:

$$X_{k+1} = X_k + hb(X_k) + \sqrt{h}\sigma(X_k)\xi_{k+1},$$

$\xi_k \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0, I)$, and $h \ll 1$. We can write it in a sde:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t.$$

4.1 IS for Markov Chains

The density for the k steps of a Markov Chain is:

$$\pi(x_1, ..., x_k) = \pi(x_1, ..., x_{k-1})\pi(x_k|x_1, ..., x_{k-1}) = \pi(x_1, ..., x_{k-1})q(x_k|x_{k-1}) = \prod_2^k q(x_k|x_{k-1})\pi(x_1).$$

Now we'll choose a reference density, where:

$$\tilde{\pi}(x_1, ..., x_k) = \tilde{\pi}(x_1)\prod_2^k \tilde{q}(x_k|x_{k-1})\tilde{\pi}(x_1).$$

So that:

$$\frac{\pi}{\tilde{\pi}}(x_1, ..., x_k) = \frac{\pi(x_1)}{\tilde{\pi}(x_1)} \prod_2^k \frac{q(x_k|x_{k-1})}{\tilde{q}(x_k|x_{k-1})}.$$

The second moment of the IS weight is:

$$\begin{aligned}
\mathbb{E}_{\tilde{\pi}}\left[\left(\frac{\pi(x_1, ..., x_k)}{\tilde{\pi}(x_1, ..., x_k)}\right)^2\right] &= \mathbb{E}_{\tilde{\pi}}\left[\left(\frac{q(x_k|x_{k-1})}{\tilde{q}(x_k|x_{k-1})}\right)^2 \left(\frac{\pi(x_1, ..., x_{k-1})}{\tilde{\pi}(x_1, ..., x_{k-1})}\right)^2\right] & (1)\\
&= \mathbb{E}_{\tilde{\pi}}\left[\mathbb{E}_{\tilde{\pi}}\left[\left(\frac{q(x_k|x_{k-1})}{\tilde{q}(x_k|x_{k-1})}\right)^2 |X_{k-1}\right] \left(\frac{\pi(x_1, ..., x_{k-1})}{\tilde{\pi}(x_1, ..., x_{k-1})}\right)^2\right] & (2)\\
&\geq \mathbb{E}_{\tilde{\pi}}\left[\left(\mathbb{E}_{\tilde{\pi}}\left[\left(\frac{q(x_k|x_{k-1})}{\tilde{q}(x_k|x_{k-1})}\right) |X_{k-1}\right]\right)^2 \left(\frac{\pi(x_1, ..., x_{k-1})}{\tilde{\pi}(x_1, ..., x_{k-1})}\right)^2\right] & (3)\\
&= \mathbb{E}_{\tilde{\pi}}\left[\left(\frac{\pi(x_1, ..., x_{k-1})}{\tilde{\pi}(x_1, ..., x_{k-1})}\right)^2\right] & (4)
\end{aligned}$$

by Jensen inequality.

So the second moment increases with k. That is the variance of an IS scheme increases (usually exponentially) with k. But for any fixed k we learnt last week that IS can be very effective.

4.2 SDE

Consider the recursion: $X_0^h = X_0$, and:

$$X_{k+1}^h = X_k^h + hb(X_k^h) + \sqrt{h}\sigma(X_k^h)\xi_{k+1},$$

where $\xi_{k+1} \sim \mathcal{N}(0, I)$. Let $h \to 0$ and recall the stochastic analysis theory:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t.$$

The solution is:

$$f(X_t) = f(X_0) + \int_0^t (f'(X_s)b(X_s) + \frac{f''(X_s)}{2}\sigma^2(X_s))\,\mathrm{d}s + \int_0^t f'(X_s)\sigma(X_s)\,\mathrm{d}B_s.$$

Here we consider the SDE:

$$\mathrm{d}X_t = (b(X_t) + \sigma(X_t)v(t, X_t))\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t.$$

The solution is:

$$\mathbb{E}[F(\tilde{X})e^{-\int_0^t v(s,\tilde{X}_s)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t \|v(s,\tilde{X}_s\|^2\,\mathrm{d}s}|X_0 = x_0] = \mathbb{E}[F(X_s)|X_0 = x_0].$$

In other words, if π is the distribution of the trajectory of the X, and $\tilde{\pi}$ is the distribution of the trajectory of the $\tilde{X}$, then:

$$\frac{\mathrm{d}\pi}{\mathrm{d}\tilde{\pi}} = e^{-\int_0^t v(s,\tilde{X}_s)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t \|v(s,\tilde{X}_s\|^2\,\mathrm{d}s}.$$

The IS estimation is that:

$$\tilde{F}_M = \frac{1}{M}\sum_{j=1}^M F(\tilde{X}^{(j)})Z_{0,T}^{(j)},$$

where $Z_{0,T}^{(j)}$ is generated by a Langevin's process.

$$Z_{r,T} = e^{-\int_r^t v(s,X_s)\,\mathrm{d}B_s - \frac{1}{2}\int_r^t \|v(s,X_s)\|^2\,\mathrm{d}s},$$

for $0 \leq r \leq t$. And

$$\frac{\mathrm{Var}\tilde{F}_M}{\mathbb{E}[F(X)]^2} = \frac{1}{M}\left(\frac{\mathbb{E}[F(\tilde{X})^2 Z_{0,T}^2]}{\mathbb{E}[F(X)]^2} - 1\right).$$

Now we want to estimate $\mathbb{E}[e^{-g(X_T^\epsilon)/\epsilon}]$ when:

$$\mathrm{d}X_t^\epsilon = b(X_t^\epsilon)\,\mathrm{d}t + \sqrt{\epsilon}\sigma(X_t^\epsilon)\,\mathrm{d}B_t$$

By the analysis of the sde, we have:

$$-\epsilon \log \mathbb{E}[e^{-g(X_T^\epsilon)/\epsilon} | X_t^\epsilon = x] = G^\epsilon(t,x) \xrightarrow{\epsilon \to 0} G(t,x) = \inf_\phi \left\{\frac{1}{2}\int_t^T \|\phi_s\|^2\,\mathrm{d}s + g(y_T^\phi)\right\},$$

where

$$y_t^\phi = x \quad , \quad \frac{\mathrm{d}}{\mathrm{d}s} y_s^\phi = b(y_s^\phi) + \sigma(y_s^\phi)\phi_s.$$

It gives the shortest path from the starting point to the finishing point.

So we have:

$$\epsilon \log \frac{\mathbb{E}[e^{-2g(X_T^\epsilon)} Z_{0,T}^2]}{\mathbb{E}[e^{-g(X_T^\epsilon)})]^2} = \inf_\phi \left\{\frac{1}{2}\int_0^T \|\phi_s\|^2\,\mathrm{d}s + 2g(y_T^\phi)\right\} - \inf_\phi \left\{\frac{1}{2}\int_0^T \|\phi_s\|^2\,\mathrm{d}s + g(y_T^\phi)\right\} \geq 0$$

which gives an exponential rate of $\frac{1}{\epsilon}$.

We have that:

$$\frac{\mathrm{Var}\tilde{F}_M}{\mathbb{E}[e^{-g(x_T/\epsilon)}]} = \frac{1}{M}\left(\frac{e^{-V^\epsilon(0,x)/\epsilon}}{e^{-2G^\epsilon(0,x)/\epsilon}} - 1\right), \tag{5}$$

and $\epsilon \log \frac{e^{-V^\epsilon(0,x)/\epsilon}}{e^{-2G^\epsilon(0,x)/\epsilon}} = 2G^\epsilon(0,x) - V^\epsilon(0,x) \xrightarrow{\epsilon \to 0} 2G(0,x) - V(0,x)$. There exists a zero-variance choice of b.

5 Exercise

Exercise 1.1

Find $\gamma(\epsilon)$ for $f(x) = x$ and $\pi = \mathrm{Exp}(1)$.

Exercise 1.2

Lower bound of $\liminf_{M\to\infty} \frac{1}{M} \log \mathbb{P}(\bar{f}_M - \pi[f] > \epsilon)$.

Exercise 6.1

Our exponential twist IS scheme from last week:

$$\tilde{\pi}(x_1, ..., x_k) = \frac{\pi(x_1, ..., x_k) e^{-w^T x}}{\mathbb{E}_\pi[e^{-w^T x}]}$$

Check that if π is Markov than $\tilde{\pi}$ is a Markov.