AE567 Statistical Inference, Estimation, and Learning
University of Michigan January 25 2023

Project 1: Stochastic Modeling, Monte Carlo methods, and Variance Reduction

Due Date: Tuesday, February 20, 2024 by 11:59 pm on Canvas

1 Instructions

The deliverable for this project is a report detailing your approach and answers to all of the questions in each section. This is an individual project: all code, answers, and analysis must be your own. However, discussions on piazza and amongst colleagues about general ideas is allowed. You may also use any code that has been covered during lecture – provided you understand what it is doing.

Formatting instructions are given below. **Note: Submissions not following the formatting instructions will not be graded.**

1. Two files must be submitted. These files are a single PDF file and a single Zip file.
2. The single PDF file contains your text, figures, and the code as an appendix. The name of the pdf file should be `<lastname>_project_1.pdf`. This file must appear first on canvas due to the way that the speed-grader works.
3. The single Zip file contains a repository of your code
4. **Code is included in two parts of the submission, as indicated above.** The code must *both* be attached in the *appendix*, and be available as part of a zip file. In the appendix, it must be legible.
5. If you would like to explicitly write code in the main writeup (i.e., not in the appendix), you may do so as long as the code is a BRIEF 5 line snippet. Longer snippets will not count as following the formatting instructions, and significant point deductions will be made.
6. The preference is that a formal typed report is submitted. However, a scanned written report that is legible will be accepted. Illegible reports will not be evaluated.
7. If you use an LLM-based chat tool (e.g., chatGPT), this counts as "code". The full transcript of any interaction must be provided in the appendix. I will view your prompts and interactions and consider that in the evaluation. Note that you are responsible for the final report and all its content, if you use the tool incorrectly, or

trust an answer that it gives that is incorrect, or you don't cite a primary source that it may quote from to support your method, then you will be penalized. *(This is an experimental requirement and may change in the future)*

8. If the above mentioned instructions are not followed an opportunity to resubmit with a 5% penalty will be given.

9. Every answer must be justified in detail, and this includes providing the necessary background. When equations are used they must either be written or referred to in the lecture notes explicitly. The steps of a calculation must be clear. Any assumptions made should be written down explicitly. As an example, some of the questions below deal with importance sampling. To answer these questions thoroughly you should recall the basic idea of importance sampling, then provide the specific choices of proposal, number of samples, etc. Also describe how you came up with that choice of proposal distribution.

10. Justifying an approach using an answer such as "Implementation is shown in the code in the appendix" will not be accepted. Describe your implemention in the text.

11. All figures/tables must have captions, they must be labeled, and they must be referenced explicitly in the text.

12. All plots must have labeled axes and any relevant legend must be shown. A caption must be included to determine what is being plotted. All text and font sizes must be legible. If values span multiple orders of magnitude, use a log scale.

13. All questions must be clearly answered, any sections of the report must be clearly delineated.

The grade distribution will be *roughly* as follows

1. Well-written and logically structured report 20%. Note: this does not mean perfect grammar or spelling (I am not concerned with the quality of your English writing), it just means that it is easy to follow and understand

2. Correct coding 10%

3. Demonstrated ability to use concepts from lectures to analyze the results, provide their meaning, and to comment on unexpected results. 70%

If your code doesn't work, that does not mean you will not get a good grade. However, you will have to do extra work in this case to demonstrate that you do understand the concept, but that you just ran into issues coding.
Example:

1. "The answer is XXX, however this is wrong because it doesnt follow theorem XXX" *This will only get partial/minimal credit*

2. "The answer is XXX, however this answer underpredicts what I expected because it doesnt follow theorem XXX. Theorem XXX requires the result to be YYY. In my code I have verified that the assumptions XX, YY, ZZ are followed. Underpredicting could be indicative that VVV is wrong. I have also tried AA, BB, CC to because they address issues DD EE and FF; however, these did not work" *This type of response will get virtually all credit*

2 Warmup: life without a CLT

Suggested deadline: February 1st

Consider a Pareto-distributed random variable X, with probability density

$$p_X(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases}.$$

Let $\alpha = 3/2$. Write a simple Monte Carlo method to estimate the mean of X, i.e., $\mu_X \equiv \mathbb{E}[X]$.

Let $\bar{x}_n$ denote your n-sample Monte Carlo estimator of μ_X. Numerically investigate the convergence of this estimator. In particular, run many (10^2 or more) independent sequences $(\bar{x}_n)_{n\in\mathbb{N}}$ and examine each sequence. Then look at the sampling distribution of the estimator at various values of n. Does there seem to be an asymptotic distribution? Or a discernable convergence rate? Provide an explanation of your numerical findings.[1]

3 Importance sampling for random walks

Suggested deadline: February 13th

Random walks have found application in countless fields, ranging from ecology and economics to physics, chemistry, biology, and robotics. They also form the basis of more complex topics such as markov chains. Here we will investigate how one can sample random walks and to compute probabilities of events.

3.1 1-D Bernoulli random walk

As a starting point, we first consider a standard random walk in one dimension. The walker starts out at the origin and can take a fixed number of steps $N = 100$. Each step is of length 1 and is randomly chosen to point either left or right with equal probability. That is,

$$S = \sum_{j=1}^{N} X_j$$

[1] If you would like to understand this behavior better, after performing your numerical experiments, please see I. Zaliapin, Y. Kagan, F. Schoenberg, "Approximating the distribution of Pareto sums," *Pure and Applied Geophysics*, **162**: 1187–1228 (2005).

where $X_1, X_2, \ldots, X_N$ are i.i.d. random variables with $\mathbb{P}(X_j = 1) = \mathbb{P}(X_j = -1) = 0.5$ for each j.

Questions 2.1
Note: For full credit on the importance sampling questions your answers should not use importance sampling to sample S directly, the sampling should occur on the X_j

(a) Write a code to simulate the N-step random walk. Make a visualization of realizations of $S_n = \sum_{j=1}^{n} X_j$ as a function of n.

(b) Use a simple Monte Carlo method with 10^5 trials to compute the probability $\mathbb{P}(S > 10)$.

(c) Use *importance sampling* with 10^5 trials to compute the probability $\mathbb{P}(S > 55)$.

(d) For comparison, derive analytical expressions for the desired probabilities.

(e) Estimate the errors in your Monte Carlo estimates of probability in parts (b) and (c):

(i) First, evaluate the Monte Carlo standard error for each estimate. Then use the central limit theorem to write an approximate 95% confidence interval for the desired probability (i.e., for $\mathbb{P}(S > 10)$ or $\mathbb{P}(S > 55)$ as appropriate).

(ii) Next, run at least 1000 independent replicates of your Monte Carlo estimators from parts (b) and (c). For each replication, compute a 95% confidence interval for the desired probability. Report how many of the confidence intervals you get actually contain the true value of the probability.

(iii) Finally, rather than considering estimates only after 10^5 trials, extract a sequence of "running mean" estimators with the number of trials M ranging from 1 to 10^5 in each sequence. Do this for each of your 1000 independent replicates. Plot the "envelope" of these parallel sequences versus M and evaluate an empirical 95% band inside the envelope. How does this band compare to that predicted by the CLT?

3.2 3-D Gaussian random walk

Next we consider a three-dimensional Gaussian random walk:

$$\mathbf{S} = \sum_{j=1}^{N} \mathbf{X}_j$$

where each $\mathbf{X}_j$ is a three-dimensional random vector composed of independent standard normals, i.e., $\mathbf{X}_j = (x_1, x_2, x_3)$ and $x_i \sim N(0, 1)$ for $i = 1, 2, 3$. This is a simple model for PMD in optical fibers.

Suppose we are interested in the probability that the total distance traveled in $N = 100$ steps is larger than a certain value L: $\mathbb{P}(|\mathbf{S}| > L)$, where $|\mathbf{S}| = \sqrt{s_1^2 + s_2^2 + s_3^2}$.

Questions 2.2

(a) Write a code to simulate the N-step random walk.

(b) Use a simple Monte Carlo method with 10^5 trials to compute the probability $\mathbb{P}(|\mathbf{S}| > 10)$.

(c) Use importance sampling with 10^5 trials to compute the probability $\mathbb{P}(|\mathbf{S}| > 55)$. Think intuitively about how to construct a good biasing distribution in this case!

(d) Estimate the errors in your Monte Carlo estimates of probability in parts (b) and (c). (Monte Carlo standard errors and/or confidence intervals are sufficient here.)

Note: in this example, it is also possible to derive an analytical expression for the desired probability and to use it to verify your simulation results.

4 Multilevel Monte Carlo and Control Variates for Stochastic ODEs

Suggested deadline: February 20th

The random walk described above has a limiting distribution called a Wiener Process. Let $X_1, X_2, \ldots,$ be i.i.d random variables with mean 0 and variance 1. Define a time-varying process according to

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} \sum_{1 \le k \le \lfloor nt \rfloor} X_k, \quad t \in [0,1]; \tag{1}$$

as $n \to \infty$ this process approaches a Wiener process $W^{(n)}(t) \to W(t)$. A Wiener process is characterized by the following properties

1. $W(0) = 0$ a.s.
2. $W(t)$ has independent increments: $W(t_1), W(t_2 - t_1), \ldots, W(t_k - t_{k-1})$ are independent for all $0 \le t_1 \le t_2 \cdots < t_k$.
3. $W(t)$ has Gaussian increments: $W(t + \Delta t) - W(t) \sim \mathcal{N}(0, \Delta t)$ – i.e., they have zero mean and variance Δt.
4. $W(t)$ has continuous paths: With probability 1, $W(t)$ is continuous in t.

This process forms the basis of stochastic modeling in dynamical systems by providing the mechanism for introducing randomness. Such models are often written in the following form

$$dY_t = b(t, Y_t)dt + h(t, Y_t)dW_t, \tag{2}$$

where b is the *drift* and h is the *diffusion*. If $\sigma = 0$ then we obtain a standard deterministic ordinary differential equation. Otherwise, we have a stochastic ordinary differential equation (SDE). Loosly speaking, this means that the process "on average" wants to evolve according to the dynamics defined by b, but there are random perturbations that affect it every

moment. These random perturbations are distributed according to a Gaussian distribution. We will make this idea more explicit when we return to state estimation later in the course.

To simulate this stochastic process we will use a numerical integration algorithm called the Euler-Maruyama scheme. This algorithm has three steps:

1. Partition $[0, 1]$ into N intervals with spacing $\Delta t = 1/N$.
2. Specify the initial condition Y_0.
3. Sequentially iterate according to the following equation

$$Y_{n+1} = Y_n + b(t_n, Y_n)\Delta t + h(t_n, Y_n)\Delta W_n, \tag{3}$$

   where $\Delta W_n = W(t_{n+1}) - W(t_n)$.

In the above, we have used the convention that each of the following are equivalent $Y_n = Y_{t_n} = Y(t_n)$ and $t_n = \Delta t\ n$.

Question:

1. What is the distribution of ΔW_n?

4.1 Geometric Brownian Motion

In this question we will look at ways to obtain the expected values of stochastic differential equations, and we will consider a special case of *Geometric Brownian Motion* where

$$b(t, Y_t) = \mu Y_t,$$
$$h(t, Y_t) = \sigma Y_t,$$

with $\mu = 0.05$, $\sigma = 0.2$, $Y_0 = 1$, and $0 < t < 1$. For instance, this can be considered as a very simple model of the price of a stock. We will also consider a payoff function for the stock that is given by the random variable

$$P = \exp(-0.05)\max(0, Y(1) - 1).$$

MLMC Questions

1. Write a code to simulate Geometric Brownian Motion for a given timestep Δt.
   - Verify the mean and variance of $Y(1)$ using the analytic solution:

$$Y_t = Y_0 \exp\left(\left(\mu - \sigma^2/2\right)t + \sigma W_t\right).$$

     You may use Monte Carlo to verify your code.

- What is the variance of a Monte Carlo estimator of the mean?

2. Adjust your code to simulate two processes, one with Δt and $4\Delta t$, that share the underlying realization of the Wiener process every $4\Delta t$.

3. Write a code to perform multilevel Monte Carlo to estimate the expectation $\mathbb{E}[Y(1)]$ and verify your analytic solution. Modify your code for estimating $\mathbb{E}[P]$.

   - Use a four-level MLMC scheme with $(\Delta t)_i = 4^{-i}$ for $i = 2, 3, 4, 5$ for each of the levels. We will call the result that has a timestep of 4^{-5} the "high fidelity" model.
   - Make a plot of the variance of each level of this estimator. Discuss.
   - Make a plot of the expectation of each level of this estimator. Discuss.

4. What is the theoretical cost of the MLMC estimator? Use this theoretical cost to determine an expression for the optimal sample allocation for a given target accuracy.

5. Using the optimal sample allocation for several target accuracies to determine

   - What would be the required number of Monte Carlo samples of the high-fidelity model to achieve the same accuracies?
   - Make a plot showing Estimator variance vs. Equivalent number of high-fidelity evaluations for both Monte Carlo and Multilevel Monte Carlo. Discuss.

4.2 Extra Credit: Solution of Deterministic Boundary Valued Problems

One fascinating aspect about stochastic differential equations is that they can be used to solve *some* deterministic PDEs. You may come across this type of duality in many application areas, but here we will look at one particular case called the *Dirichlet-Poisson Boundary Value problem*. I will provide you with enough background to solve this problem; however, for more detailed information as to **why** this works please see e.g., *Stochastic Differential Equations: An introduction with applications* by Oksendal.
Let us consider the Poisson equation in one dimension with Dirichlet boundary conditions, i.e.,

$$\frac{\partial^2 u(x)}{\partial x^2} = -g(x), \quad x \in (0,1) \tag{4}$$

$$u(0) = u(1) = 0 \tag{5}$$

In words, the second derivative of the solution $u(x)$ is equal to $-g(x)$ in the interior of a domain (defined on the unit interval). On the boundaries of the interval, the solution is zero $u(0) = u(1) = 0$. As an example if $g(x) = -\sin(5x)$ then the solution

$$u(x) = \frac{\sin(5)}{25}x - \frac{\sin(5x)}{25} \tag{6}$$

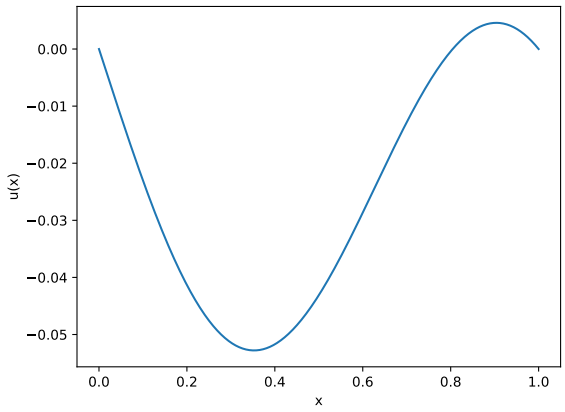

Figure 1: Solution of Equation (4)

is visualized in Figure 1.
Amazingly, this solution can also be represented as an expectation of and SDE with $b(t, Y) = 0$ and $h(t, Y) = \sqrt{2}$. In particular, we can write

$$u(x) = \mathbb{E}\left[\int_0^\tau g(Y(t))dt\right] = \mathbb{E}\left[\phi(x)\right], \tag{7}$$

where τ is the time at which $Y(t)$ exists the boundary $(0, 1)$. In other words, to sample the integral within the expectation (equivalently to sample $\phi(x)$) you must follow the following three steps:

1. Initialize $Y_0 = x$.

2. Simulate Y_t until $Y_t > 1$ or $Y_t < 0$. Set τ to be the t at which this happens and n to be the number of timesteps.

3. Approximate the integral as $\int_0^\tau g(Y(t))dt \approx \sum_{i=0}^n g(Y_{t_i})\Delta t$

This procedure essentially refers to the fact that if you want the solution at $u(x)$, then start a random walk from x and wait until you hit the boundary. Along the way accumulate a "cost" that is incurred, and that is defined by the right-hand-side of the PDE.

Note: The above three steps are used to obtain a single sample of $\phi(x)$. A MC procedure would then solve this many times to obtain multiple samples to accurately evaluate the expectation.

MLMC Questions: Let $x = 0.5$

- Does your multilevel Monte Carlo approach work for this problem? Justify your answer with the same arguments as for the previous question.
- Do you have any other ideas for reducing the variance of a Monte Carlo estimate of the solution?