Stochastic Simulation
Professor: Prof. Fabio Nobile
Assistant: Eliott Van Dieren

Week of 2025-10-06 to 2025-10-10

Lab 5 of Thursday 9th October 2025

Exercice 1.

A simulator would like to produce an unbiased estimate of $\mathbb{E}(XY)$, where the two independent random variables X and Y have bounded first moments and can be generated by a stochastic simulation. To this end, they simulate $R \in \mathbb{N}$ replications $X_1, \dots, X_R$ of X and, independently of this, R replications $Y_1, \dots, Y_R$ of Y. They thus have the following two natural estimators for $\mathbb{E}(XY)$ at their disposal:

$$\mathrm{Est}_1 := \left(\frac{1}{R}\sum_{r=1}^{R} X_r\right)\left(\frac{1}{R}\sum_{r=1}^{R} Y_r\right) \quad \text{and} \quad \mathrm{Est}_2 := \frac{1}{R}\sum_{r=1}^{R} X_r Y_r \,.$$

1) Verify that both estimators Est_1 and Est_2 are unbiased.

2) Show that $\mathrm{Var}(\mathrm{Est}_1) < \mathrm{Var}(\mathrm{Est}_2)$.

3) Use the delta method to show that $\sqrt{R}(\mathrm{Est}_1 - \mu_x\mu_y) \xrightarrow{d} N(0, \tau^2)$. Find τ^2 explicitly and derive a $1-\alpha$ asymptotic confidence interval.

Exercice 2.

We are interested in estimating the expectation $\mathbb{E}[X]$ of a random variable X using several sampling methods. We will work with three different distributions of X.

- $X \sim \mathrm{Pareto}(x_m = 1, \gamma = 3.1)$ (i.e. with PDF $p(y) = \mathbb{1}_{y > x_m} x_m^{\gamma} \gamma y^{-(\gamma+1)}$), $\mathbb{E}[X] = \frac{\gamma x_m}{\gamma - 1}$ for $\gamma > 1$, else $\mathbb{E}[X] = \infty$ and $\mathrm{Var}(X) = \frac{x_m^2 \gamma}{(\gamma-1)^2(\gamma-2)}$ for $\gamma > 2$, else $\mathrm{Var}(X) = \infty$. Lastly, the normalised third absolute moment is $\beta = \mathbb{E}[|X - \mathbb{E}[X]|^3] / \mathrm{Var}(X)^{3/2}$ for $\gamma > 3$ For $\gamma = 3.1, x_m = 1$, we have $\mathbb{E}[X] = \frac{3.1}{2.1}$, $\mathrm{Var}(X) = \frac{3.1}{2.1^2 \cdot 1.1}$ and $\beta \approx 18.9$

- $X \sim \mathrm{Lognormal}(\mu = 0, \sigma = 1)$, $\mathbb{E}[X] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$, $\mathrm{Var}(X) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$, and
$$\beta = \frac{4\,\mathrm{Erf}(\frac{\sigma}{2\sqrt{2}}) - 3e^{\sigma^2}\,\mathrm{Erf}(\frac{3\sigma}{2\sqrt{2}}) + e^{3\sigma^2}\,\mathrm{Erf}(\frac{5\sigma}{2\sqrt{2}})}{(e^{\sigma^2} - 1)^{3/2}}.$$
For $\mu = 0, \sigma = 1$, we have $\mathbb{E}[X] = \sqrt{e}$, $\mathrm{Var}(X) = e(e-1)$ and $\beta \approx 6.35$.

- $X \sim U([-1, 1])$, $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = \frac{1}{3}$, and $\beta = 3\sqrt{3}/4$.

We will start with a Crude Monte Carlo estimator

$$\overline{I}_N := \frac{1}{N}\sum_{k=1}^{N} X_k,$$

for $X_1, X_2, \dots, X_N$ i.i.d. samples of X.

1) By *a priori* analysis, determine $N(\alpha, \epsilon)$, $\alpha = 0.05$ such that

$$\mathbb{P}(|\overline{I}_N - \mathbb{E}[X]| > \epsilon) < \alpha$$

using Chebycheff's inequality, the Berry-Esseen Theorem and the CLT

$$\frac{\overline{I}_N - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)/N}} \sim N(0,1).$$

   a) Provide the analytical expressions for $N(\alpha, \epsilon)$ using the three different techniques. Furthermore, compute $N(\alpha, \epsilon)$ for $\epsilon = 0.1$ and $\alpha = 0.05$ for each of the three distributions above, and compare the obtained values for the three techniques. What is the order of growth of $N(\alpha = 0.05, \epsilon)$ w.r.t ϵ?

   b) Generate M independent MC approximations $\overline{I}_N^j$, $j = 1, 2, \dots, M$ with $M = 1000$. Compute empirically $\mathbb{P}(|\overline{I}_N - \mathbb{E}[X]| > \epsilon)$, and compare with the target $\alpha = 0.05$.

2) Algorithm 1 proposes a sequential Monte Carlo method to compute the expectation $\mathbb{E}[X]$ of a random variable X, where the sample size is doubled at each iteration until the estimated $1 - \alpha$ confidence interval based on a central limit theorem approximation is smaller than a prescribed tolerance ϵ. The algorithm then outputs the final sample size $N(\epsilon, \alpha)$, as well as the estimated value $\bar{X}_N$.

Algorithm 1: Sample Variance Based SMC

Input: N_0, distribution λ, accuracy $\epsilon > 0$, confidence $1 - \alpha > 0$.
Output: $\overline{X}_{\epsilon,\alpha}$ (i.e, approximation of $\mathbb{E}[X]$ with $X \sim \lambda$), N.

Set $k = 0$, generate N_k i.i.d. replica $\{X_i\}_{i=1}^{N_k}$ of $X \sim \lambda$ and

$$\bar{X}_{N_k} = \frac{1}{N_k}\sum_{i=1}^{N_k} X_i, \tag{2.1}$$

$$\overline{\sigma}^2_{N_k} := \frac{1}{N_k - 1}\sum_{i=1}^{N_k}(X_i - \overline{X}_{N_k})^2. \tag{2.2}$$

while $\bar{\sigma}_{N_k} C_{1-\alpha/2}/\sqrt{N_k} > \epsilon$ **do**
  Set $k = k + 1$ and $N_k = 2N_{k-1}$.
  Generate a new batch of N_k i.i.d. replicas $\{X_i\}_{i=1}^{N_k}$ of $X \sim \lambda$.
  Compute the sample variance $\overline{\sigma}^2_{N_k}$ by (2).
end while
Set $N = N_k$, generate i.i.d. samples $\{X_i\}_{i=1}^{N}$ of λ and compute the output sample mean $\overline{X}_{\epsilon,\alpha}$.

Algorithm 1 can be particularly sensitive to the choice of initial sample size N_0, and as such, we would like to assess the robustness of such an algorithm in estimating $\mathbb{E}[X]$ for different distributions of X. For some values of N_0 ranging between 10 and 50, consider $\alpha = 10^{-1.5}$ and $\epsilon = 1/10$.

a) Repeat the simulation $K = 20\alpha^{-1}$ times and record the sample sizes $\{N^{(i)}\}_{i=1}^{K}$ as well as the computed sample means $\{\bar{X}_{\epsilon,\alpha}^{(i)}\}_{i=1}^{K}$ returned by the algorithm for each run $i = 1, ..., K$.

b) Estimate the probability of failure $\overline{p}$ of the algorithm :

$$\overline{p}_K(N_0, \epsilon, \alpha) = \frac{1}{K} \sum_{i=1}^{K} \mathbb{1}_{|\bar{X}_{\epsilon,\alpha}^{(i)} - \mathbb{E}[X]| > \epsilon}.$$

Then check whether $\overline{p}_K(N_0, \epsilon, \alpha) \leqslant \alpha$ holds.

c) Repeat your experiment for different values of ϵ and α. Discuss your results.

Hint. You may generate Pareto(x_m, α) r.v. by inversion.

3) One drawback of Algorithm 1 is that it requires to generate a batch of samples at each iteration, which may be computationally expensive. Compare Algorithm 1 with the sequential Monte Carlo method in Algorithm 2, where one realization is added at a time. Do you observe any difference in the performance of the two algorithms ? Is one more robust numerically than the other?

Algorithm 2: One-at-a-time Sample Variance Based SMC

Input: N_0, distribution λ, accuracy $\epsilon > 0$, confidence $1 - \alpha > 0$.

Output: $\overline{X}_{\epsilon,\alpha}$ (i.e, approximation of $\mathbb{E}[X]$ with $X \sim \lambda$), N.

Set $k = 0$, generate N_k i.i.d. samples $\{X_i\}_{i=1}^{N_k}$ of λ and compute the sample variance

$$\overline{\sigma}_{N_k}^2 := \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (X_i - \overline{X}_{N_k})^2. \tag{2.3}$$

while $\bar{\sigma}_{N_k} C_{1-\alpha/2} / \sqrt{N_k} > \epsilon$ **do**

Set $k = k + 1$ and $N_k = N_{k-1} + 1$.

Generate a new i.i.d. sample $X^{(N_k+1)}$ of λ.

Compute

$$\bar{\mu}_{N_k+1} = \frac{N_k}{N_k + 1} \bar{\mu} + \frac{1}{N_k + 1} X^{(N_k+1)} \tag{2.4}$$

$$\bar{\sigma}_{N_k+1}^2 = \frac{N_k - 1}{N_k} \sigma_{N_k}^2 + \frac{1}{N_k + 1} (X^{(N_k+1)} - \bar{\mu}_{N_k})^2 \tag{2.5}$$

end while

Set $N = N_k$, generate i.i.d. samples $\{X_i\}_{i=1}^{N}$ of λ and compute the output sample mean $\overline{X}_{\epsilon,\alpha}$.

---

4) Repeat the points above by replacing the CLT stopping criterion in Algorithm 1 and Algorithm 2 by the Chebycheff and Berry-Esseen based stopping criteria.

Exercice 3.

Consider the problem of pricing a Barrier option with maturity $T > 0$ based on the stock price S, which is given as the solution to the stochastic differential equation

$$dS = rS\,dt + \sigma S\,dW\,, \quad S(0) = S_0\,,$$

where W denotes a standard one-dimensional Wiener process. One can show that $S_t = S_0 e^{X_t}$, where $X_t = (r - \sigma^2/2)t + \sigma W_t$ with W being a standard Wiener process. It follows that S_t has a log-normal distribution for any $t > 0$. For $m \in \mathbb{N}$, let $t_i = i\varDelta t$ with $\varDelta t = T/m$ denote the discrete observation times of the stock price S (e.g. daily at market closure). The payoff of a call option subject to a lower barrier is then given by

$$\varPsi(S_{t_0}, S_{t_1}, \dots, S_T) = (S_T - K)_+ \mathbb{I}_{\left\{B \leqslant \min_{i=0,\dots,m}(S_{t_i})\right\}}\,,$$

where $B < S_0$ denotes the Barrier and $K \leqslant S_0$ the strike price. Here, $z_+ = (|z| + z)/2$ denotes the positive part of z. Estimate the expected payoff $\mathbb{E}(\varPsi(S_{t_0}, S_{t_1}, \dots, S_T))$ with antithetic variables, using the process parameters $m = 1000$, $r = 0.5$, $\sigma = 0.3$, $T = 2$, $S_0 = 5$, and $K = 10$. Specifically, investigate the variance reduction effect for different barrier values B.