Practicals

stanislasduche@gmail.com

1 Simulations of Random Variables

R Reminder. `runif` allows simulating *i.i.d.* realizations from the distribution $\mathscr{U}([0,1])$. R provides random generators for most common distributions. However, they will not be used in this section except for comparison purposes.

1.1 Inverse Function and Transformations

Exercise 1 *(Simulation of a discrete random variable).* Let X be a discrete random variable over the set $\{5,6,7,8\}$. The distribution ν of X is defined by

$$\mathbb{P}[X=5]=0.4, \quad \mathbb{P}[X=6]=0.2, \quad \mathbb{P}[X=7]=0.3, \quad \mathbb{P}[X=8]=0.1$$

1. Simulate a sample **x** of 10000 i.i.d. random variables following the distribution ν using the inverse function method.
2. Compare the barplot of the sample **x** to that of v.

R Reminder. `barplot (height = ...)` plots the barplot of categorical or discrete variables. height is a vector containing the height of the bars or the contingency table of the sample. For a sample **x**, the contingency table can be obtained with `table(x)`.

Self-Evaluation. Number of times the following elements appear in the solution.

Question	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
1.	0	0	0	0	0	0

Exercise 2 *(Exponential distribution and related distributions).* Let $X_1,\ldots,X_n$ be i.i.d. random variables following the exponential distribution $\mathscr{E}(\lambda)$, i.e. $\mathbb{E}[X_1]=\lambda^{-1}$.

1. (a) Simulate 10000 realizations of the distribution $\mathscr{E}(\lambda)$ for $\lambda=2$ using the inverse function method.
(b) Check with a histogram and a Quantile-Quantile plot that the distribution of this sample matches the distribution $\mathscr{E}(\lambda)$.
2. Recall that $S_n=X_1+\ldots+X_n$ follows the gamma distribution $\Gamma(n,\lambda)$, i.e. $\mathbb{E}[S_n]=n\lambda^{-1}$.
(a) Based on this result, simulate 10000 realizations of the gamma distribution $\Gamma(n,\lambda)$ with $\lambda=1.5$ and $n=10$.
(b) Graphically verify that the distribution of this sample matches the gamma distribution $\Gamma(n,\lambda)$.
3. Let $N=\sup\{n\geq 1 : S_n\leq 1\}$ (by convention $N=0$ if $S_1>1$). Then N follows the Poisson distribution $\mathscr{P}(\lambda)$.
(a) Based on this result, simulate 10000 realizations of the Poisson distribution $\mathscr{P}(\lambda)$ with $\lambda=4$.
(b) Graphically verify that the distribution of this sample matches the Poisson distribution $\mathscr{P}(\lambda)$.

R Reminder

- `hist(x, freq = F)` displays the histogram of a sample x. The freq option specifies whether the histogram is represented in frequency density (`freq = TRUE` by default) or probability density (`freq = FALSE`).
- `lines`(x, y) adds a piecewise linear curve connecting the points with abscissa `x` and ordinate `y`.
- `quantile(`x`, probs)` returns the quantiles of a sample x for a vector of probabilities `probs`.
- `dexp, pexp`, and `qexp` correspond respectively to the density, distribution function, and quantile function of an exponential distribution.
- `dgamma, pgamma`, and `qgamma` correspond respectively to the density, distribution function, and quantile function of a gamma distribution.
- `dpois, ppois`, and `qpois` correspond respectively to the density, distribution function, and quantile function of a Poisson distribution.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
1.	0	0	0	0	0	0
2.	0	≤ 1	0	0	0	0
3.	0	≤ 1	1	0	0	0

1.2 Normal Distribution, Gaussian Vectors, and Brownian Motion

Exercise 3 *(Box-Muller Algorithm).*

1. Write a function BM(n) that returns n realizations of the normal distribution $\mathcal{N}(0,1)$ using the Cartesian version of the Box-Muller method.
2. Validate the algorithm using a graphical tool.

R Reminder. `dnorm`, `pnorm`, and `qnorm` correspond respectively to the density, distribution function, and quantile function of a normal distribution.

Self-Evaluation. Number of times the following elements appear in the solution.

Question	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
1.	0	0	0	≤ 1	0	0

Exercise 4 *(Simulation of Gaussian Vectors).* Let $\mathbf{X} = (X_1, X_2)$ follow the distribution $\mathcal{N}(\mu, \Sigma)$, with

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$$

1. Simulate a sequence of vectors $\left(X^{(n)}\right)_{n\geq 1} = (X_{1,n}, X_{2,n})_{n\geq 1}$ that follow the distribution of $\mathbf{X}$.
2. What is the distribution of $X_1 + X_2$? Validate this result graphically.

Self-Evaluation. Number of times the following elements appear in the solution.

Question	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
1.	0	0	0	0	0	0

Exercise 5 *(Simulation of Brownian Motion).* Using the properties of Brownian increments, simulate a realization of Brownian motion at the times $(t_1, \ldots, t_{1110})$ defined by $t_i = i/100$ for $i \in [\![1, 100]\!]$, $t_i = 1 + (i - 100)/10$ for $i \in [\![101, 110]\!]$, and $t_i = 2 + (i - 110)/1000$ for $i \in [\![111, 1110]\!]$.

Self-Evaluation. Number of times the following elements appear in the solution.

c()	for	while	if-else	Vectorize	apply
0	0	0	0	0	0

1.3 Rejection Algorithm

Exercise 6 *(Rejection - A First Example).* Let f be a density function defined for all real numbers x by

$$f(x) = \frac{2}{\pi}\sqrt{1 - x^2}\mathbb{1}_{\{x \in [-1,1]\}}$$

1. Use the rejection method to simulate 10,000 realizations following the distribution with density f.
2. Plot the histogram of this sample and compare it to the density f.

Self-Evaluation. Number of times the following elements appear in the solution.

Version	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
$\star$ | $\star\star\star$	0 | 1	1 | 0	1	0	0	0

Exercise 7. Use the rejection method to simulate 5,000 realizations following the uniform distribution over the unit disk $\mathscr{D} = \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\right\}$.

Self-Evaluation. Number of times the following elements appear in the solution.

Version	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
$\star$ | $\star\star\star$	0 | 1	1 | 0	1	0	0	0

Exercise 8 *(Truncated Normal Distribution).* The normal distribution $\mathcal{N}\left(\mu, \sigma^2\right), \mu > 0, \sigma > 0$, truncated with support $[b, +\infty[$ has a density f defined for all real numbers x by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}\Phi\left(\frac{\mu - b}{\sigma}\right)} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\mathbb{1}_{\{x \geq b\}}$$

where Φ is the cumulative distribution function of the standard normal distribution $\mathscr{N}(0, 1)$.
Instrumental Density n°1: $\mathcal{N}\left(\mu, \sigma^2\right)$.

1. Write a function `tr_norm (n, b, mean, sd)` to simulate n realizations following the normal distribution $\mathcal{N}\left(\texttt{mean}, \texttt{sd}^2\right)$ truncated with support $[b, +\infty[$.
2. Simulate 10,000 realizations following the normal distribution $\mathscr{N}(0, 2)$ truncated with support $[2, +\infty[$. Validate your algorithm graphically.

Instrumental Density n° 2. The translated exponential distribution from $b, \tau\mathscr{E}(\lambda, b)$, has a density

$$g_\lambda(x) = \lambda e^{-\lambda(x-b)} \mathbb{1}_{\{x \geq b\}}, \quad x \in \mathbb{R}.$$

In the following, we fix λ to the optimal value obtained in TD°2.

3. Write a function `tr_norm_2 (n, b, mean, sd)` to simulate n realizations following the normal distribution $\mathscr{N}\left(\texttt{mean}, \texttt{sd}^2\right)$ truncated with support $[b, +\infty[$.

4. Simulate 10,000 realizations following the normal distribution $\mathscr{N}(0, 2)$ truncated with support $[2, +\infty[$. Validate your algorithm graphically.

5. Compare the performances of `tr_norm` and `tr_norm_2`.

Self-Evaluation. Number of times the following elements appear in the solution.

Version	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
1.⋆ | ⋆⋆⋆	0 | 1	1 | 0	1	0	0	0
4.⋆ | ⋆⋆⋆	0 | 1	1 | 0	1	0	0	0

2 Classical Monte Carlo Method

Exercise 9 *(Estimation of π).*

1. Using the classical Monte Carlo method with $n = 150000$ draws, propose an estimation of π based on the generation of
(a) $U_1, \dots, U_n$ i.i.d. random variables with uniform distribution $\mathscr{U}([0,1])$;
(b) $(U_{1,1}, U_{2,1}), \dots, (U_{1,n}, U_{2,n})$ pairs of i.i.d. random variables following the distribution $\mathscr{U}([0,1]) \otimes \mathscr{U}([0,1])$.

2. Which estimator is the most efficient?

3. Using Bienaymé-Tchebychev's inequality or the asymptotic regime hypothesis at a 95% confidence level, for what value of n do we achieve a precision of 10^{-2}? Discuss the results obtained.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	c( )	for	while	if-else	Vectorize	apply
1.	0	0	0	0	0	0
3.	0	0	0	0	0	0

Exercise 10 *(Integral Calculation).* We are interested in calculating the following integral

$$\delta = \int_2^{+\infty} \int_0^5 \sqrt{x+y} \sin\left(y^4\right) \exp\left(-\frac{3x}{2} - \frac{y^2}{4}\right) \mathrm{d}x \, \mathrm{d}y$$

1. Propose an estimation of δ using the classical Monte Carlo method with:
(a) a uniform distribution generator and a normal distribution generator;
(b) an exponential distribution generator and a normal distribution generator;
(c) an exponential distribution generator and a truncated normal distribution generator.

2. Determine the computational cost of these methods. Conclude which method is preferable to use.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	c ( )	for	while	if-else	Vectorize	apply
1.$\star$ | $\star\star\star$	0 | 1	1 | 0	1	0	0	0

3 Variance Reduction Methods

3.1 Importance Sampling

Exercise 11 *(Cauchy Distribution).* Let X be a random variable following the Cauchy distribution $\mathscr{C}(0,1)$. We are interested in estimating

$$p := \mathbb{P}[X \geq 50] = \int_{50}^{+\infty} \frac{1}{\pi\left(1+x^2\right)} \mathrm{d}x.$$

We propose to compare the classical Monte Carlo method and the importance sampling method based on the Pareto distribution whose cumulative distribution function is given for $x \in \mathbb{R}$ by

$$F(x) = \left[1 - \left(\frac{a}{x}\right)^k\right] \mathbb{1}_{\{x \geq a\}}, \quad \text{with} \quad a > 0 \quad \text{and} \quad k > 0.$$

We will consider $n = 10000$ draws.

1. Provide an estimation of p from simulations following the Cauchy distribution $\mathscr{C}(0,1)$.
2. (a) What values of a and k should be chosen for the importance sampling method?
   (b) Use these to estimate p with the importance sampling method.
3. Determine the relative efficiency between these two methods.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	c( )	for	while	if-else	Vectorize	apply
2.	0	0	0	0	0	0
3.	0	0	0	0	0	0

Exercise 12 *(Rejection vs. Importance Sampling).* We revisit the calculation of the integral

$$\delta = \int_2^{+\infty} \int_0^5 \sqrt{x+y} \sin\left(y^4\right) \exp\left(-\frac{3x}{2} - \frac{y^2}{4}\right) \mathrm{d}x \ \mathrm{d}y$$

1. By expressing δ as an expectation with respect to the exponential distribution and the translated exponential distribution, propose an estimation of δ using the importance sampling method.
2. Compare the performance of this method with the classical Monte Carlo method based on an exponential distribution generator and a truncated normal distribution generator.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	c ( )	for	while	if-else	Vectorize	apply
2.$\star$ | $\star\star\star$	0 | 1	1 | 0	1	0	0	0

3.2 Other Variance Reduction Methods

Exercise 13 *(Gaussian Integral and Variance Reduction).* We aim to estimate

$$\mathscr{I} = \int_0^2 e^{-x^2}\, \mathrm{d}x$$

using Monte Carlo methods. In this exercise, we will consider $n = 10000$ draws.

1. Provide an estimation of $\mathscr{I}$ using the classical Monte Carlo method based on two different distributions.
2. For these estimators, propose a new estimation of $\mathscr{I}$ based on an antithetic variable.
3. (a) Calculate the k^{th} moment for $k \in \mathbb{N}^*$ of a uniform random variable on $[0, 2]$.
   (b) Use the series expansion of $x \mapsto \exp\left(-x^2\right)$ truncated at order $k \in \mathbb{N}^*$ as a control variate. Propose a method to find the value of k to use in practice.
   (c) Deduce an estimation of $\mathscr{I}$ using the control variate method.
4. Consider an estimator based on the $\mathscr{U}([0, 2])$ distribution generator.
   (a) Choose a stratification variable Z and a partition $D_1, \ldots, D_K$ such that $\mathbb{P}[Z \in D_k] = 1/K$.
   (b) Provide an estimation of $\mathscr{I}$ using the stratified sampling method with proportional allocation and $K = 10$. Study the influence of K on the precision of the method.
5. Determine the relative efficiency of these different methods.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	c ()	for	while	if-else	Vectorize	apply
3.b.	0	2	0	0	0	0
3.c.	0	1	0	0	0	0
4.b. ($\star$)	0	1	0	0	0	0
4.b. ($\star\star\star$)	0	0	0	0	0	1
4.c.	0	1	0	0	0	0

Exercise 14 *(Brownian Motion and Finance).* Let $\mathbf{X} = (X_1, X_2)$ be a Gaussian vector with distribution $\mathscr{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

We want to estimate

$$\mathscr{I} = \mathbb{E}\left[\max\left\{0, \frac{1}{2}e^{-\sigma^2/2+\sigma X_1} + \frac{1}{2}e^{-\sigma^2/2+\sigma X_2} - 1\right\}\right].$$

For numerical applications, we will use $\sigma = 2$, $n = 5000$ draws, and provide a 95% asymptotic confidence interval.

1. Provide an estimation of $\mathscr{I}$ using the classical Monte Carlo method.

We are initially interested in the antithetic variate method. Let

$$\mathbf{Z} \sim \mathscr{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \quad \text{and} \quad A = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

2. (a) Show that the vector $A\mathbf{Z}$ has the same distribution as $\mathbf{Z}$.
   (b) Provide an estimation of $\mathscr{I}$ using the antithetic variate method.
   (c) Does the antithetic variate method reduce variance?

Finally, we consider the control variate method.

3. (a) Calculate $\mathbb{E}\left[\exp\left(\sigma X_1\right) + \exp\left(\sigma X_2\right)\right]$.
   (b) Deduce an estimation of $\mathscr{I}$ using the control variate method.

4. Determine the relative efficiency of these three methods.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	c( )	for	while	if-else	Vectorize	apply
1.	0	0	0	0	0	0
2.	0	0	0	0	0	0
3.	0	1	0	0	0	0

Exercise 15. The number of precipitation events S in a month is modeled by a Poisson distribution with parameter $\lambda = 3.7$. The amount of water Q_s falling during a precipitation event s is modeled by a Weibull distribution with shape parameter $k = 0.5$ and scale parameter $\lambda = 2$ (assuming that the precipitation events are independent). The total amount of rain in a month is

$$X = \begin{cases} 0 & \text{, if } S = 0 \\ \sum_{s=1}^{S} Q_s & \text{, otherwise} \end{cases}$$

We are interested in months with low precipitation and aim to estimate $p = \mathbb{P}[X < 3]$ (i.e., less than 3 cm of rain per month).

1. Provide an estimation of p using the classical Monte Carlo method. Provide the 95% confidence interval.

2. Provide an estimation of p using the stratified sampling method with proportional allocation, specifying the strata used. Provide the 95% confidence interval.

3. Propose a method to estimate p using the stratified sampling method with optimal allocation. What difficulties do you encounter?

4. Compare the variances and relative efficiency of these three estimation methods. Discuss the results briefly.

Self-Evaluation. Number of times the following elements appear in the solution.

Questions	`c( )`	`for`	`while`	`if-else`	`Vectorize`	`apply`
1.⋆ | ⋆⋆⋆	0	1 | 0	0	0	0	0 | 1
2.⋆ | ⋆⋆⋆	0	5 | 2	0	0	0	0 | 4