Second Lab Session: introduction to Monte Carlo methods

UE Computational Statistics

moi

> ⚠ Attention
>
> Remember to put your name in the header of the document.

In this session, we will

- simulate random numbers by inverting the cumulative distribution function,
- reduce the variance of a Monte Carlo estimator by using different methods,
- compare different methods with confidence intervals for the mean.

```python
import numpy as np
import scipy
import scipy.stats as stats
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Part 1: Inverting the cumulative distribution function

1.1 The Weibull distribution

The Weibull distribution $\mathrm{Wei}(\lambda, k)$ is a continuous probability distribution with two parameters:

- $\lambda > 0$ is the scale parameter,
- $k > 0$ is the shape parameter.

It is often to model the lifetime of objects, that is to say time to failure. In this context, the value of k is critical: if $k < 1$, the failure risk decreases over time, if $k = 1$, the failure risk is constant over time, and if $k > 1$, the failure risk increases over time.

Its probability density function is given, for all $x \in \mathbb{R}$, by

$$w(x|\lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \mathbf{1}\{x \geq 0\}.$$

Exercise:

1. Plot the probability density functions of the Weibull distribution with $\lambda = 1$ and $k = 0.5, 1, 2, 5, \ldots$ in the same graph. According to you, what is happening when k increases?
2. Show that, when $X \sim \mathrm{Wei}(\lambda, k)$, then $X/\lambda \sim \mathrm{Wei}(1, k)$
3. Compute the cumulative distribution function $W(x|\lambda, k)$ of the Weibull distribution explicitly.
4. Compute the quantile function $W^{-1}(p|\lambda, k)$ of the Weibull distribution explicitly.
5. Show that, the expected value of $X \sim \mathrm{Wei}(\lambda, k)$ is

$$\mathbb{E}(X) = \lambda\Gamma\left(1 + \frac{1}{k}\right), \quad \text{where } \Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)dt.$$

   Actually $\Gamma(x)$ is Euler's gamma function. It is implemented in the `scipy.special` module.

1.2 Basic Monte Carlo simulations

Exercise:

1. Write a function `inverse_cdf` that takes as input (1) p a probability, (2) λ and k the parameters of the Weibull distribution and returns the quantile.
2. Simule a large sample $x_0, x_1, \ldots, x_N$ of size $N = 1{,}000$ from the Weibull distribution with $\lambda = 1$ and $k = 2$ using the inverse method.
3. Plot

$$\bar{x}_n = \frac{x_0 + \cdots + x_{n-1}}{n}$$

   as a function of n for $n = 1, 2, \ldots, N$. Add an horizontal line at the expected value of X, using the formula of the previous exercise. What is your conclusion?

1.3 Confidence intervals

An asymptotic confidence interval for the mean of a random variable X is given by

$$\bar{x}_n \pm z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}},$$

where $z_{1-\alpha/2}$ is the quantile of order $1 - \alpha/2$ of the standard normal distribution, and σ is the standard deviation of X. If σ is not available in closed form (which is often the case), we can use the empirical standard deviation

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n-1}(x_i - \bar{x}_n)^2}.$$

In fact, there is a high chance that, when we resort to Monte Carlo to compute $\mathbb{E}(X)$ numerically, $\mathbb{V}(X)$ is also unknown.

In the case of the Weibull distribution, we can compute the variance of X explicitly. Indeed, we have

$$\mathbb{V}(X) = \lambda^2 \left(\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right).$$

Exercise: 1. Using the previous simulation, compute the empirical standard deviation of $x_0, \dots, x_{n-1}$ and both bounds of the asymptotic confidence intervals around $\bar{x}_n$, for $n = 2, \dots, N$ 2. Add the confidence intervals to the previous plot. What is your conclusion?

Part 2: Reducing the variance

2.1 Antithetic variates

Assume that Q is the quantile function of a distribution with CDF F. Assume that Q is an increasing function over $(0;1)$. If U is a random variable uniformly distributed on $[0,1]$, then $Q(U)$ is distributed according to F. This is the principle of the inverse method. On the other hand, $1-U$ is also uniformly distributed on $[0,1]$. Therefore, $Q(1-U)$ is also distributed according to F. Using both $Q(U)$ and $Q(1-U)$ is the principle of the antithetic variates method.

Exercise:

1. Compare $X = Q(U)$, $Y = Q(1-U)$ and $Q(1/2)$.
2. Look at Chebyshev's sum inequality and show that X and Y are negatively correlated.
3. Show that $Z = (X+Y)/2$ has the same expected value than X, but a smaller variance than X and Y.

2.2 Antithetic variates applied to the Weibull distribution

Exercise: Apply the antithetic variates method to the Weibull distribution with $\lambda = 1$ and $k = 2$. Compare the asymptotic confidence intervals given by the sample of X and the sample of Z, with graphs. What is your conclusion?

2.3 Control variates

Assume that X is a random variable and Y is a random variable such that $\mathbb{E}(Y)$ is known. The control variates method consists in estimating $\mathbb{E}(X)$ by

$$\bar{x}_n - \beta(\bar{y}_n - \mathbb{E}(Y)),$$

where $\bar{x}_n$ and $\bar{y}_n$ are the sample means of X and Y respectively, and β is the solution of

$$\beta = \frac{\mathbb{V}(X,Y)}{\mathbb{V}(Y)}.$$

Often, the value of β is unknown. But we can estimate it by

$$\hat{\beta} = \frac{\sum_{i=0}^{n-1}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=0}^{n-1}(y_i - \bar{y}_n)^2}$$

and use it in the control variates method.

Exercise:

1. Compute the variance of
$$Z_n = \bar{X}_n - \lambda(\bar{Y}_n - \mathbb{E}(Y))$$
when $(X_0, Y_0), \dots, (X_{n-1}, Y_{n-1})$ are i.i.d. samples of a joint distribution whose components can be correlated.
2. Show that the variance is minimal for $\lambda = \beta$
3. What is happening when X and Y are independent? Is the method useful?

2.4 Control variates applied to the Weibull distribution

Exercise:

1. Prove that, when $k = 1$, $\mathrm{Wei}(\lambda, 1)$ is the exponential distribution with parameter λ.
2. The expected value of the exponential distribution is explicit and easy to compute. Compute it explicitly.
3. Draw a large sample of $(Q(U|\lambda, k), Q(U|\lambda, 1))$ and apply the control variates method to estimate the expected value of $\mathrm{Wei}(\lambda, k)$.
4. Compare the asymptotic confidence intervals given by the sample of X and the sample of Z, with graphs. What is your conclusion?
5. Can we combine the antithetic variates and the control variates methods?