3 Monte Carlo Methods

Deterministic problems: Monte Carlo techniques can be used to evaluate mathematical expressions, namely, integrals, systems of equations or more complicated mathematical models. Sometimes standard approximations from numerical analyses can be used but in some cases only Monte Carlo methods could provide a solution.

Problems with randomness: A Monte Carlo study can be used to approximate the distribution of some variable or statistics, or to address some specific properties of the statistics.

3.1 Parameter estimation

One of the applications of the Monte Carlo method is to estimate some characteristics of random variables.

Many characteristics of the random variables are defined through integrals:

For example:

- $\mathrm{E}(X) = \int_D x f(x) dx;$
- $\mathrm{E}(g(X)) = \int_D g(x) f(x) dx;$
- $\mathrm{Var}(X) = \int_D (x - \mathrm{E}(x))^2 f(x) dx;$
- $\mathrm{P}(X < a) = \int_{-\infty}^{a} f(x) dx.$

These integrals do not always have an analytical solution!

We can use Monte Carlo methods to estimate these integrals.

Let

$$\boxed{\theta = \int_D h(x)dx.}$$

If we can decompose the function h as the product of two functions with one of them being the density function of a random variable with support D, that is,

$$\boxed{h(x) = g(x)f(x)}$$

with

$$\int_D f(x)dx = 1, \quad (D \text{ is the support of the random variable})$$

and $f(x) \geq 0$, then we can interpret the parameter θ as an expected value,

$$\boxed{\theta = \int_D h(x)dx = \int_D g(x)f(x)dx = \mathrm{E}(g(X)).}$$

If we have a random sample $x_1, x_2, ..., x_n$ from the random variable X with density f, an estimator of θ is:

$$\boxed{\hat{\theta} = \frac{\sum_{i=1}^{n} g\left(x_i\right)}{n}.}$$

Law of large numbers: If $X_1, X_2, ..., X_n$ is a sequence of independent and identically distributed random variables then,

$$\boxed{\frac{\sum_{i=1}^{n} g\left(X_i\right)}{n} \rightarrow \mathrm{E}\left(g(X)\right).}$$

almost sure when n converges to infinite.

Let X be a random variable with exponential distribution, $X \sim Exp(3)$.
Our goal is to estimate the expected value and variance of the random variable $g(X)$, where $g(x) = e^{-x^2}, \quad x \geq 0$.

$$\theta = E(g(X)) = \int_0^{+\infty} e^{-x^2} 3e^{-3x} dx;$$

$$\upsilon = Var(g(X)) = \int_0^{+\infty} \left(e^{-x^2} - \theta\right)^2 3e^{-3x} dx.$$

These integrals do not have an analytical solution then we are going to estimate them using the Monte Carlo method.

We generate a sample of the random variable X, $x_1, x_2, ..., x_m$ and then the estimators for θ and υ are:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} e^{x_i^2};$$

$$\hat{\upsilon} = \frac{1}{n} \sum_{i=1}^{n} \left(e^{x_i^2} - \hat{\theta} \right)^2.$$

Suppose we want to evaluate $\theta = \int_0^1 g(x)dx$.

If $U \sim U(0,1)$ then $\theta = E[g(U)]$.

Then a good estimator is:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} g(u_i).$$

where $u_1, u_2, ..., u_m$ is a sample of the random variable $U \sim U(0,1)$

Remark: To evaluate $\int_a^b g(x)dx$ we change the variable $y = \dfrac{x-a}{b-a}$

Estime the value of π using the Monte Carlo method:

$$\pi = 4 \int_0^1 \sqrt{1 - x^2} dx$$

then, $\pi = E(g(U))$ where $g(u) = \sqrt{1 - u^2}$.

We generate a sample of the random variable U, $u_1, u_2, ..., u_m$ and then we can consider the estimator for π is:

$$\hat{\pi} = \frac{4}{n} \sum_{i=1}^{n} \sqrt{1 - u_i^2}.$$

The multidimensional case:

If

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 g(x_1, ..., x_n) dx_1 \cdots dx_n$$

then

$$\theta = E[g(U_1, ..., U_n)] \simeq \frac{1}{n} \sum_{i=1}^{n} g(u_1^i, ..., u_n^i).$$

Another Monte Carlo estimator for π :

$$\pi = 4\int_0^1\int_0^{\sqrt{1-x^2}} 1dydx = \int_0^1\int_0^1 4\mathrm{I}_{\{x^2+y^2\leq 1\}}(x,y)dydx = \int_0^1\int_0^1 h(x,y)dydx$$

with $h(x,y) = 4\mathrm{I}_{\{x^2+y^2\leq 1\}}(x,y)$.

Then an alternative estimator for π is:

$$\hat{\pi} = \frac{4}{n}\sum_{i=1}^{n} \mathrm{I}_{\{x^2+y^2\leq 1\}}(u_{1i}, u_{2i})$$

where

$$((u_{11}, u_{21}), (u_{11}, u_{21}), ..., (u_{1n}, u_{2n}))$$

is a sample from the uniform bidimensional random variable (U_1, U_2).

3.2 Properties of the estimators

Proposition

If $X_1, X_2, ..., X_n$ is a sequence of independent and identically distributed random variables with density f then the Monte Carlo estimator

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

of

$$\theta = E(X) = \int_D g(x)f(x)dx$$

satisfies,

$E(\hat{\theta}) = \theta$. *(Is an unbiased estimator).*

$$Var(\hat{\theta}) = \frac{1}{n}\Big[\int_D g^2(x)f(x)dx - \theta^2\Big] = \frac{1}{n}\Big[\int_D (g(x)-\theta)^2 f(x)dx\Big].$$

Proof:

- $$E\left[\frac{\sum g(X_i)}{n}\right] = \frac{1}{n}\sum E\left(g(X_i)\right) = \frac{1}{n}\sum E\left(g(X)\right) = E(X).$$

- $$Var\left[\frac{\sum g(X_i)}{n}\right] = \frac{1}{n^2}\sum Var\left(g(X_i)\right) = \frac{1}{n^2}\sum Var\left(g(X)\right) =$$
$$= \frac{1}{n}Var(g(X)) = \frac{1}{n}\left[\int_D (g(x)-\theta)^2 f(x)dx\right].$$

- An estimate based in the Monte Carlo method uses a random sample and is therefore a random variable thus it is important to have some measure of variability.

- Whenever reporting Monte Carlo estimates, it is important to report the size of the simulation and also a standard error of the estimate.

- Often we will want to run simulations until the variability of an estimator is less than some threshold value.

The variance of the Monte Carlo estimator is given by:

$$Var(\hat{\theta}) = \frac{1}{n}\Big[\int_D (g(x) - \theta)^2 f(x)dx\Big]$$

In general this integral does not have an analytical solution then we estimate it also with a Monte Carlo estimator:

Monte Carlo estimator for the Standard Deviation:

$$SE(\hat{\theta}) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(g(x_i) - \hat{\theta})^2}.$$

where $x_1, x_2, ..., x_n$ is a random sample of X.

3.2 Hypothesis Testing

Definition

(Hypothesis Testing). A statistical hypothesis is some conjecture about the distribution of one or more random variables. For each hypothesis designated by null hypothesis and denoted by H_0, there is always an alternative hypothesis denoted by H_1.

Informally:

- A statistical hypothesis is some statement about the parameters of one or more populations (parametric tests) or about the distribution of the population (non-parametric tests).

- The goal of a test is to use the information of a data sample to decide (reject or no reject) about a conjecture over unknown aspects of a given population.

There are always some risk associated to statistical inference:

Errors types:

- Type I error: reject H_0 when H_0 is true (rejecting error),
- Type II error: no reject H_0 when H_0 is false (no rejecting error).

We denote by,

$\alpha = P(\text{erro do tipo I}) = P(\text{Rejeitar } H_0 | H_0 \text{is true})$, where α is called significance level of the test. In general, we assign a very small value to the probability of type I error (0.05 ou 0.01)

$\beta = P(\text{erro do tipo } II) = P(\text{no-reject} H_0 | H_0 \text{is false})$, where $1 - \beta$ is called power of the test)

- Bilateral test:

$$H_0 : \mu = \mu_0;$$
$$H_1 : \mu \neq \mu_0.$$

- Right unilateral test:

$$H_0 : \mu \leq \mu_0;$$
$$H_1 : \mu > \mu_0.$$

- Left unilateral test:

$$H_0 : \mu \geq \mu_0;$$
$$H_1 : \mu < \mu_0.$$

Procedure to make a test using the $p-$value

- Identify the parameter of interest and fix H_0 and H_1.
- Choose the test statistic T and compute the value of the statistic with the available sample, $t_{\text{obs}} = T(x_1, ..., x_n)$.
- Compute the p-value for the computed value of the statistic:
  - Bilateral test:

$$\text{p-value} = \begin{cases} 2P(T < t_{\text{obs}}|H_0) & \text{if} \quad t_{\text{obs}} \text{ is small} \\ 2P(T > t_{\text{obs}}|H_0) & \text{if} \quad t_{\text{obs}} \text{ is high} \end{cases}$$

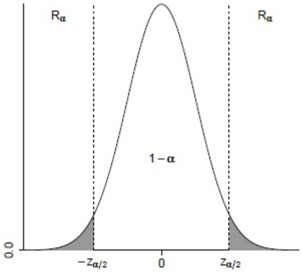
  - Right unilateral test: p-value $= P(T > t_{\text{obs}}|H_0)$
  - Left unilateral test: p-value $= P(T < t_{\text{obs}}|H_0)$

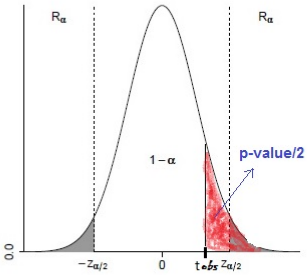
  t_{obs} small (big) means that t_{obs} is smaller (bigger) then the value specified by H_0.

- Decision: We reject H_0 if the p-value $\leq \alpha$ (significance level).

$$R_\alpha =]-\infty, -z_{\alpha/2}[\cup]z_{\alpha/2}, +\infty[,$$

$$P(T > z_{\alpha/2}) = \alpha/2 \quad \text{e} \quad P(T < -z_{\alpha/2}) = \alpha/2$$





In this case we do not reject H_0.

Suppose that $(X_1, X_2, ..., X_n)$ is a sample from a population which follows a Normal distribution with known variance and our aim is to test:

$$H_0 \,:\, \mu = \mu_0 \ \text{ vs },\, H_1 : \mu \neq \mu_0 \text{ (bilateral test)}$$

The estimator $\bar{X}$ is an unbiased estimator of μ. It is also known that $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, even if, μ, is unknown. Let us consider the following test statistic:

$$T(X_1, X_2, ..., X_n) = Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}.$$

$$\text{p-value } = \begin{cases} 2P(Z < z_{\text{obs}}|H_0) & \text{if} \quad z_{\text{obs}} \text{ is small} \\ 2P(Z > z_{\text{obs}}|H_0) & \text{if} \quad z_{\text{obs}} \text{ is large} \end{cases}$$

We know how to compute the $p-$value because we know the distribution of the statistic test Z, under the null hyphotesis.

The **empirical distribution function** is a fundamental element to make inference using the Monte Carlo method.

Given a sample of a random variable, $x_1, x_2, ..., x_n$. The empirical distribution function associated to this simple is the distribution function whose probability mass function assigns the same weight to all the elements of the sample, that is,

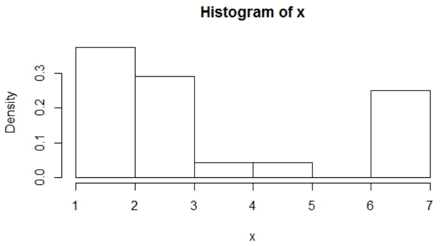
$$\boxed{\hat{F}(x) = \frac{\#\{\, i \,:\, x_i \leq x\}}{n}.}$$

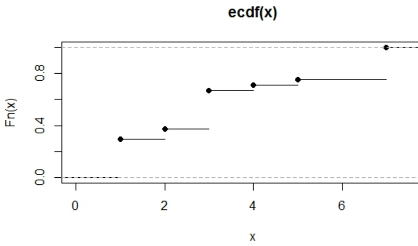
Empirical distribution function - Example

Sample=c(1,1,1,1,1,1,1,2,2,3,3,3,3,3,3,3,4,5,7,7,7,7,7,7).
(24 observations)

hist(amostra, probability = TRUE) - Draws the probability mass function graph.

plot(ecdf(x)) - Draws the empirical distribution function graph.





$P(X \leq 3) = \hat{F}(3) = p(1) + p(2) + p(3) = \frac{7}{24} + \frac{2}{24} + \frac{7}{24} = 0.6(6).$

Hypothesis testing - Example

We have a sample of independent observations of a random variable that we know follows an exponencial distribution with parameter λ. (λ is unknown.)

0,2	1,2	2,9	1,2	0,1	0,1	0,4	0,1	0,7	0,1	0,9	0,3	0,6	0,1	0,2	0,1	0,4	0,1	0,3	1,4

Our aim is to test if $\lambda = 3$.

The maximum likelihood estimator for λ,

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

Test statistic: Likelihood Ratio:

$$T(x) = \frac{L(\hat{\lambda}, x)}{L(\lambda = 3, x)}.$$

Ingredients to undertake hypothesis testing:

- Sample, $x_1, x_2, ..., x_n$;
- Null hypothesis, $H_0 : \lambda = 3$ and alternative hypothesis, $H_1 : \lambda \neq 3$;
- Significance level, α;
- Test statistic, $T(x) = \dfrac{L(\hat{\lambda}, x)}{L(\lambda = 3, x)}$;
- Distribution function of the test statistic under the null hypothesis, F_{H_0} to compute the p-value
  In this case what we know about the distribution of the test statistic is:
  $$\ln(T) \xrightarrow{d} \chi^2_{(1)}.$$
  That is, we only know the asymptotic distributions of the test statistics.

  When the sample is small we do not know the distribution of the statistic. It is a fact that there are many test statistics with **unknown distributions**.

We compute the value of the test statistic using the sample $t_{\text{obs}} = T(x)$ and then we **estimate the p-value.**

$$p\text{-value} = P(T > t_{\text{obs}} | H_0).$$

Since the distribution of T, under the null hypothesis $H_0 : \lambda = 3$, is unknown, we use an approximate distribution, the empirical distribution function, $\hat{F}_{H_0}$. For it we need to built a sample of the random variable T, $t_1, t_2, ..., t_m$. We generate m samples of the exponential random variable with parameter $\lambda = 3$, and for each sample we compute the test statistic.

$$\begin{array}{llllll}
\text{sample 1:} & x_1^1 & x_2^1 & \cdots & x_n^1 & \rightarrow & t_1 = T(x^1) \\
\text{sample 2:} & x_1^2 & x_2^2 & \cdots & x_n^2 & \rightarrow & t_2 = T(x^2) \\
 & & & \vdots & & & \\
\text{sample m:} & x_1^m & x_2^m & \cdots & x_n^m & \rightarrow & t_m = T(x^m)
\end{array}$$

The p-value is estimated using the empirical distribution function associated to the sample $t_1, t_2, ..., t_m$.

Then,

$$\hat{p}\text{-value} = \hat{P}(T > t_{t_{\text{obs}}} | H_0) = 1 - \hat{F}_{H_0}(t_{t_{\text{obs}}}) =$$

$$\boxed{= \frac{\#\{\, t_j \; : \; t_j \geq t_{\text{obs}}\} + 1}{m + 1}.}$$

Variability of the method

Let $Y = \#\{x | T(x) \geq t_0\}$

$Y \sim \text{bin}(m, p)$ where $p = \Pr(T \geq t_0)$.

Then since $\hat{p} = \dfrac{Y+1}{m+1}$,

$$E(\hat{p}) = \frac{mp+1}{m+1} \rightarrow p \text{ as } m \rightarrow \infty$$

$$\text{Var}(\hat{p}) = \frac{mp(1-p)}{(m+1)^2} \rightarrow 0 \text{ as } m \rightarrow \infty$$

The simple form of the **standard error estimator** for the variability of the method

$$SE(\hat{p}) \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{m}}$$

We remark that the standard error decreases when the number of simulated samples increases.

Statistic $T(X)$ for the previous example:

Dimension of the data set: $n = 20$

$$T(x) = \frac{L(x; \hat{\lambda})}{L(x; \lambda = 3)} = \frac{1}{(3\bar{x})^n} e^{n(3\bar{x}-1)}$$