On the behavior of the angle between two vectors in $\mathbb{R}^n$ as $n \to +\infty$ whose components follow a normal or uniform distribution

Françoise Chatelin[1] and Nahid Emad[2]

[1]Qualitative Computing Group, Cerfacs and University of Toulouse 1 Capitole, France (chatelin@cerfacs.fr)

[2]Laboratoire LI-PaRAD, Maison de la Simulation, UVSQ, Université Paris-Saclay, 78035 Versailles, France (nahid.emad@uvsq.fr)

Résumé

The purpose of this document is to study, experimentally, the behavior of the basic operations on vectors whose components are random numbers with different distribution laws. Specifically, we compute the dot product of two vectors whose components are random numbers with distribution law taken to be normal on the one hand and uniform on the other.

1 Oppenheim-Ricoux (OR) conjecture

Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be two vectors whose components x_i (resp. y_i), $i = 1, n$ are random variables with standard normal or uniform distribution. The angle between these vectors is defined by

$$\theta_n(x, y) = \measuredangle \left(\frac{x}{|| x ||}, \frac{y}{|| y ||} \right) \tag{1}$$

which is equivalent to

$$\cos(\theta_n) = \left(\frac{< x, y >}{|| x || \, || y ||} \right) \in [-1, +1] \tag{2}$$

The Oppenheim-Ricoux conjecture presented in 2014 [1] states that :

$$(OR) \qquad \lim_{n \to +\infty} | \theta_n | = \frac{\pi}{2} \Leftrightarrow \lim_{n \to +\infty} | \cos(\theta_n) | = 0 \tag{3}$$

The present note is an experimental validation of (OR).

2 Description of experiments

In order to present these experiments, we consider two real vectors x and y of length n which the components are randomly drawn with the normal and uniform distributions. We study numerically and statistically the convergence towards zero of $| \cos(\theta_n) |$ as n tends to $+\infty$. For that we compute $| \cos(\theta_n) |$ by (2) for $n \in [1, N]$ where $N \in \mathbb{N}^*$ takes increasingly large values. The statistical software that we use is R language and its environment for statistical computing and graphics [2].

2.1 Normal distribution

Here, we suppose that the components of x and y are randomly drawn with the normal distribution whose mean $\mu = 0$, variance $\sigma^2 = 1$. Two types of experiments are considered. The first type is numerical : the size N of the vectors varies. In the second type is statistical, we consider p samples of a fixed size n, $n = 10$ and $n = 1000$.

2.1.1 Normal distribution : first case

In this set of experiments, N takes 10, 100, 1000, 10000 and 100000 values successively. The left side of the figures 1-5 represents the evolution of $|\cos(\theta_n)|$ when n goes from 1 to N. We notice that for n large $|\cos(\theta_n)|$ tends to zero.

Let $z \in \mathbb{R}^n$ be the vector whose components are $z_i = \cos(\theta_i)$, $i = 1, n$. The right side of the figures 1-5 represent the distribution of the variables z_i, $i = 1, n$. The distribution seems to follow the normal law when n is large.

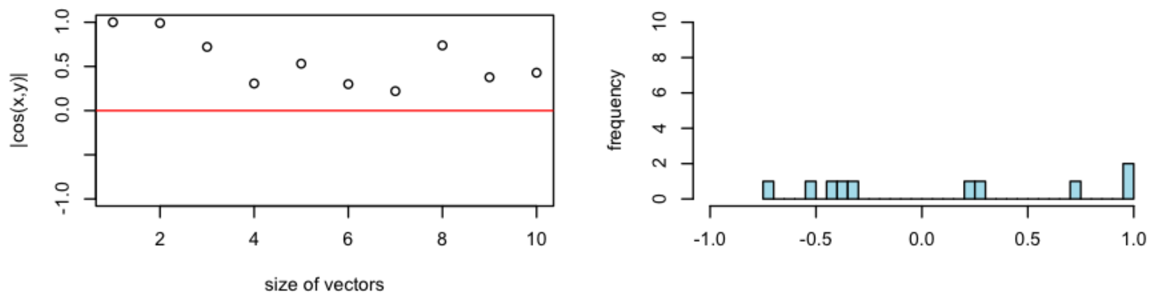

Figure 1 – $N = 1$ to 10

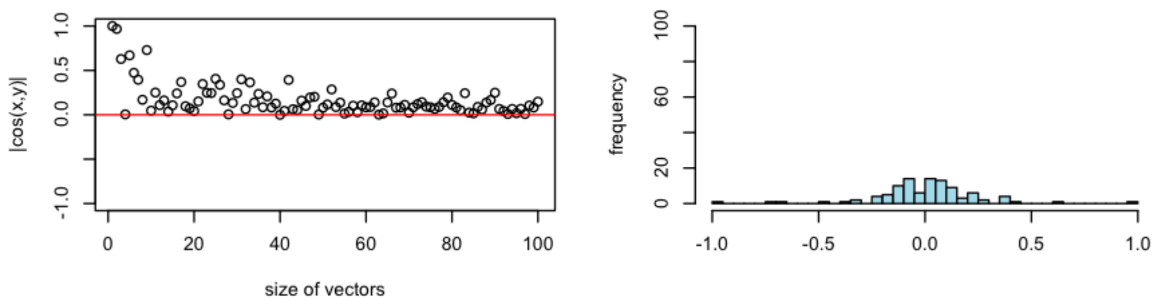

Figure 2 – $N = 1$ to 100

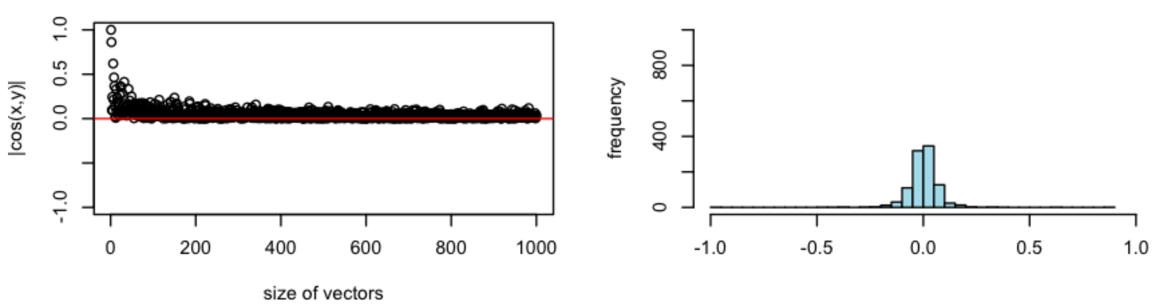

Figure 3 – $N = 1$ to 1000

Moreover, to present the evolution of $|\cos(\theta_n)|$ for growing values of n, we also show its numerical value for selected n. For that, we consider the interval $[1, N]$ divided in 10 equal segments $[\ell_{i-1}, \ell_i]$ for $i = 1, 10$ with $\ell_i = \ell_{i-1} + (N/10)$ and $\ell_0 = 0$. Then, we display the values of $|\cos(\theta)|$ on the terminals of these sub-intervals. Table 1 presents the values of $c(\ell_i) = |\cos(\theta_{\ell_i})|$ for $i = 1, 10$ for N values given in the first column.

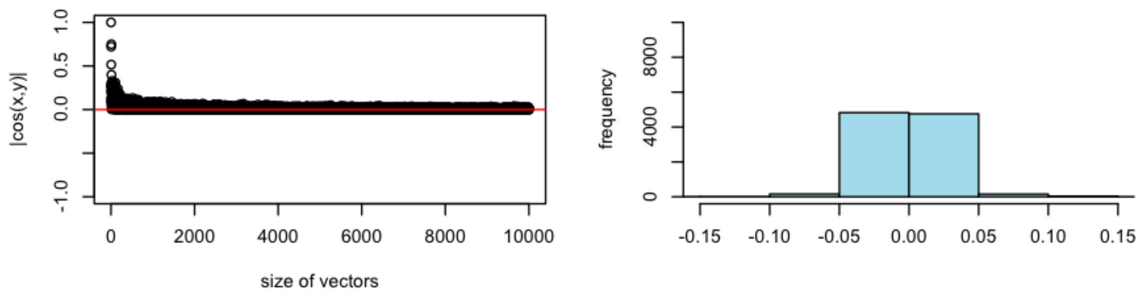

FIGURE 4 – $N = 1$ to 10000

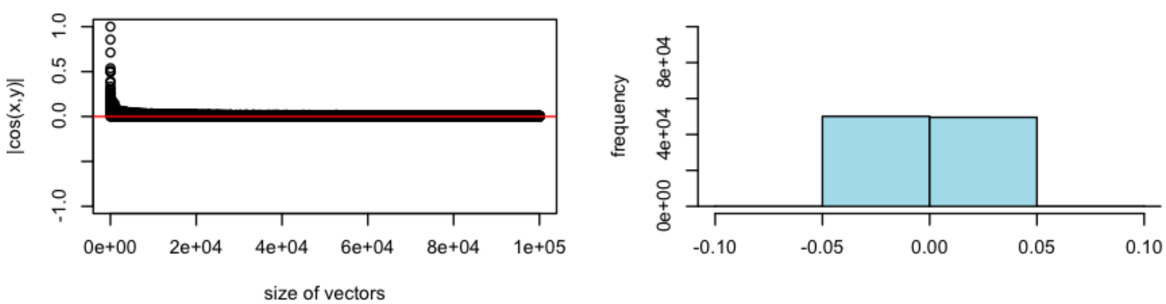

FIGURE 5 – $N = 1$ to 100000

2.1.2 Normal distribution : second case

Suppose that the components of vectors x and y of length n are randomly drawn with standard normal distribution. Let X_j and Y_j represent these random variables corresponding to the sample j and Z_j represents $\cos(\theta_n)$ for the same sample j (for fixed n). Here we want to study the mean value $\mathcal{E}_p = \frac{1}{p} \sum_{j=1}^{p} \mid Z_j \mid$ where p is the size of the sample. The left side of the figures 6-10 (resp. 11-15) represents the evolution of $\mathcal{E}_p$ when N is fixed to 10 (resp. 1000) and the size of sample varies between 1 and p (with p = 10, 100, 1000, 10000 and 100000). The right side of the figures 6-10 (resp. 11-15) represents the density of "random variables" $\mathcal{E}_i$, $i = 1, p$. We notice that this distribution seems to follow the normal law.

This table 2 shows the decay of $\mathcal{E}_p$ as a function of the growth of p. Each line of the this table presents the mean value of $\mathcal{E}_p$ for $p \in \{10, 100, 1000, 10000, 100000\}$ for a fixed n (10 for the first line and 1000 for the second line). We can notice that $\mathcal{E}_p$ tends to zero when p becomes large.

N	$c(\ell_2)$	$c(\ell_3)$	$c(\ell_4)$	$c(\ell_5)$	$c(\ell_6)$	$c(\ell_7)$	$c(\ell_8)$	$c(\ell_9)$	$c(\ell_{10})$
10	0.9902856	0.7209751	0.3066497	0.5312624	0.2996368	0.2203304	0.7378387	0.3774181	0.4285758
10^2	0.2501916	0.1506945	0.3999998	0.04415448	0.1163811	0.08798113	0.08454565	0.08260819	0.06563538
10^3	0.06487468	0.01782859	0.03096038	0.002885614	0.03940472	0.006627125	0.01246893	0.01039661	0.02977677
10^4	0.01204089	0.01838448	0.0105312	0.01846281	0.02694352	0.01389076	0.0131747	0.008603132	0.00567131
10^5	0.01455886	0.004676293	0.006317398	0.001420822	0.0002394812	0.003782211	0.006602555	0.003156112	0.00142408

TABLE 1 – $\mid \cos(\theta_n) \mid$ for $n = \ell_i$, $i = 1$ to 10 in the interval $[1, N]$.

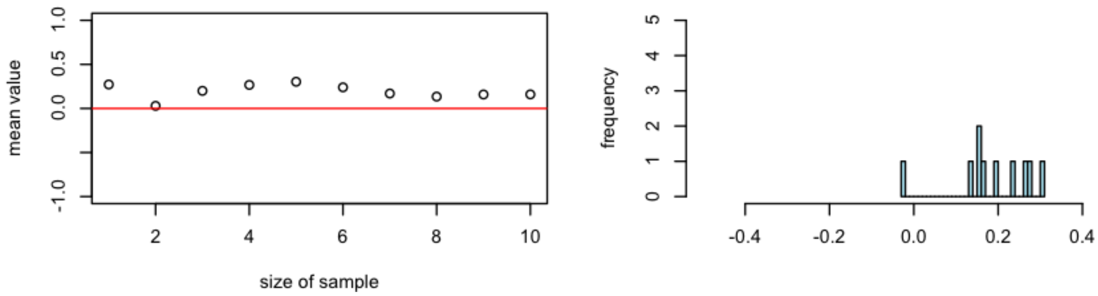

FIGURE 6 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10$ (fixed $n = 10$)

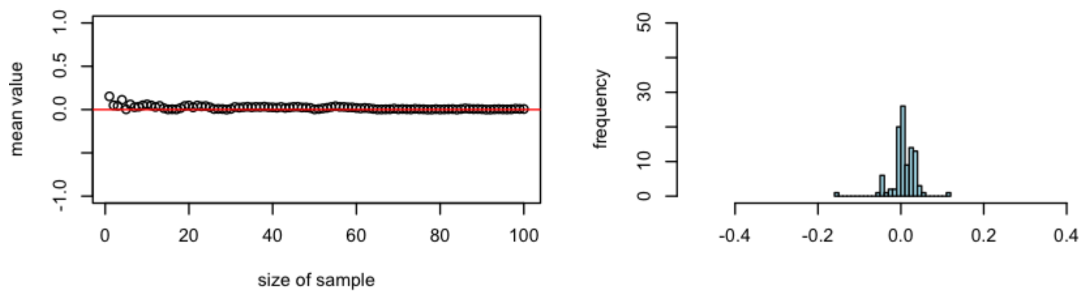

FIGURE 7 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100$ (fixed $n = 10$)

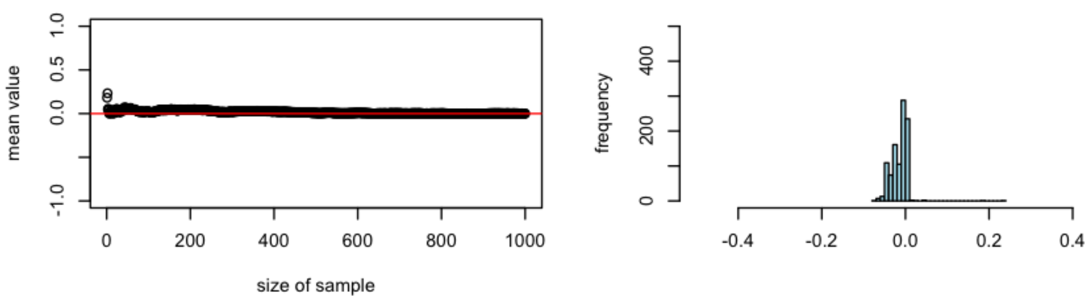

FIGURE 8 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 1000$ (fixed $n = 10$)

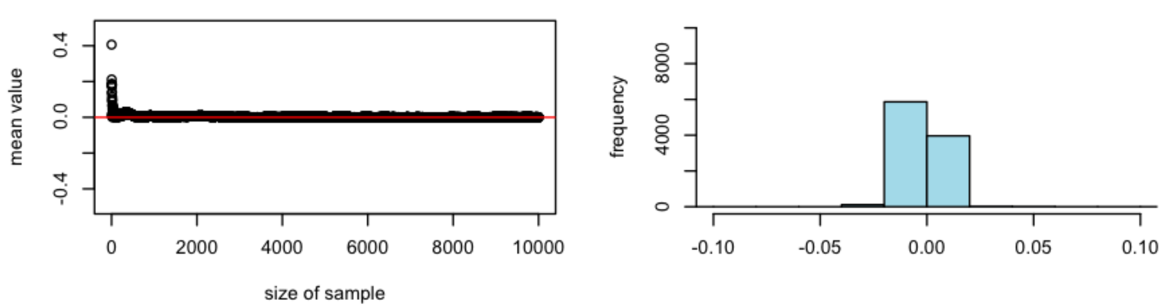

FIGURE 9 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10000$ (fixed $n = 10$)

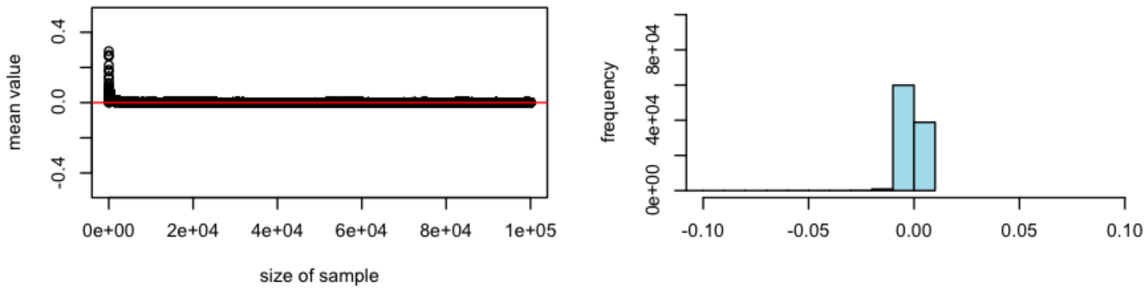

FIGURE 10 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100000$ (fixed $n = 10$)

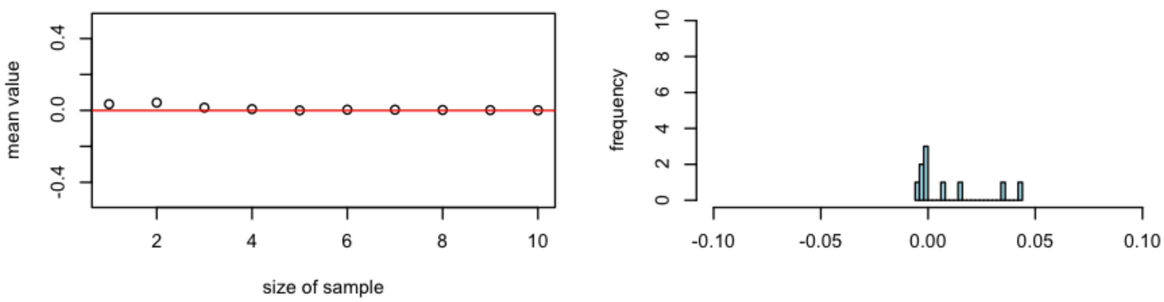

FIGURE 11 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10$ (fixed $n = 1000$)

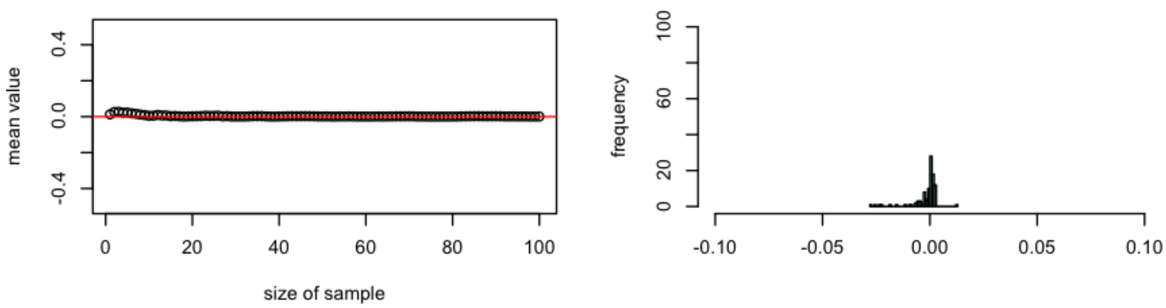

FIGURE 12 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100$ (fixed $n = 1000$)

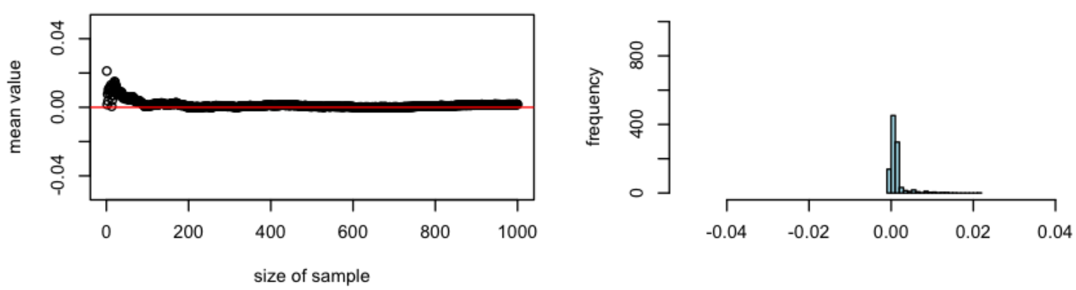

FIGURE 13 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 1000$ (fixed $n = 1000$)

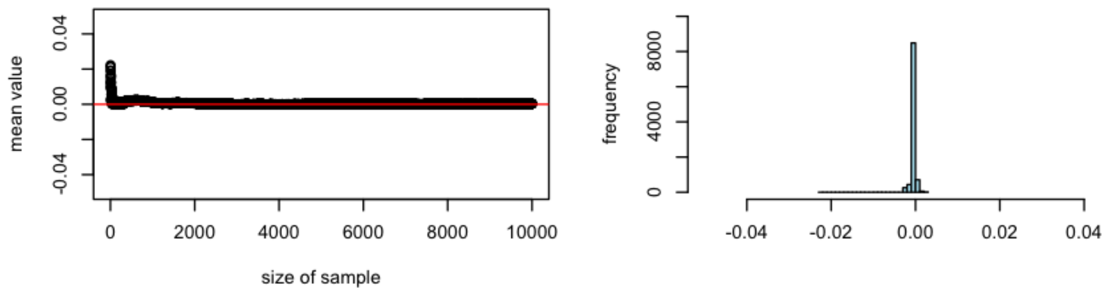

FIGURE 14 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10000$ (fixed $n = 1000$)

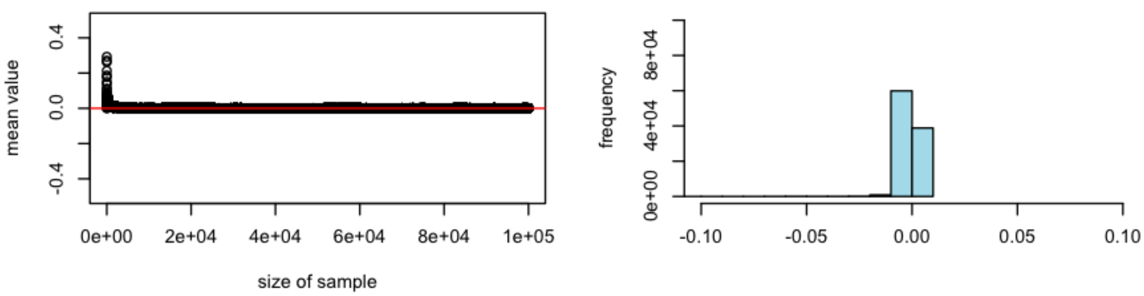

FIGURE 15 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100000$ (fixed $n = 1000$)

2.2 Uniform distribution

Here, we suppose that the the components of x and y are randomly drawn with uniform distribution on the interval $I = [a, b]$. As in section 2.1, we consider wo types of experiments. In the first one, the size of the vectors varies. In the second type, we consider p samples and the vectors of a fixed size n.

2.2.1 Uniform distribution : first case

In the set of experiments presented here N takes successively the values 10, 100, 1000, 10000 and 100000. The left side of the figures 16-20 represents the evolution of $\mid \cos(\theta_n) \mid$ when n goes from 1 to N. We notice that for n large $\mid \cos(\theta_n) \mid$ tends to zero.

Let $z \in \mathbb{R}^n$ be the vector whose components are $z_i = \cos(\theta_i)$, $i = 1, n$. The right side of the figures 16-20 represents the density of "random variables" z_i, $i = 1, n$. According to these figures, the distribution seems to follow the *normal law* when n is large.

As in the case of normal distribution law, we also show the evolution of $\mid \cos(\theta_n) \mid$ by presenting its numerical values for growing values of n. The same manner as in Section 2.1.1, we consider the interval $[1, N]$ divided in 10 equal segments $[\ell_{i-1}, \ell_i]$ for $i = 1, 10$ with $\ell_i = \ell_{i-1} + (N/10)$ and $\ell_0 = 0$. Then, we display the values of $\mid \cos(\theta) \mid$ on the terminals of these sub-intervals. Table 3 presents the values of $c(\ell_i) = \mid \cos(\theta_{\ell_i}) \mid$ for $i = 1, 10$ and N given in the first column.

n \ p	10	100	1000	10000	100000
10	0.1595175	0.006914514	0.0001481639	0.0007234779	0.0005579374
1000	0.0009650849	0.0007422512	0.00172796	0.0003205711	0.0001605052

TABLE 2 – $\mathcal{E}_p$ for various n and p

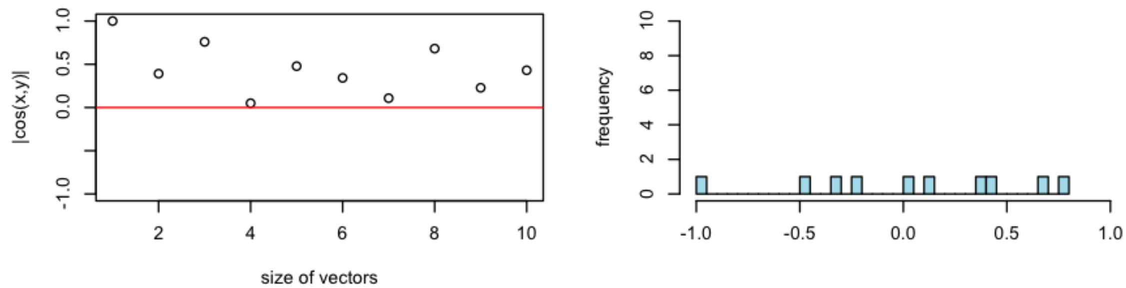

FIGURE 16 – The size of the vectors goes from 1 to 10

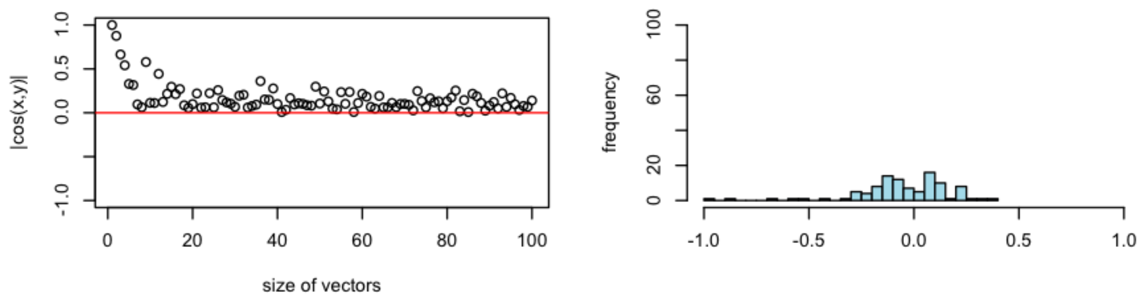

FIGURE 17 – The size of the vectors goes from 1 to 100

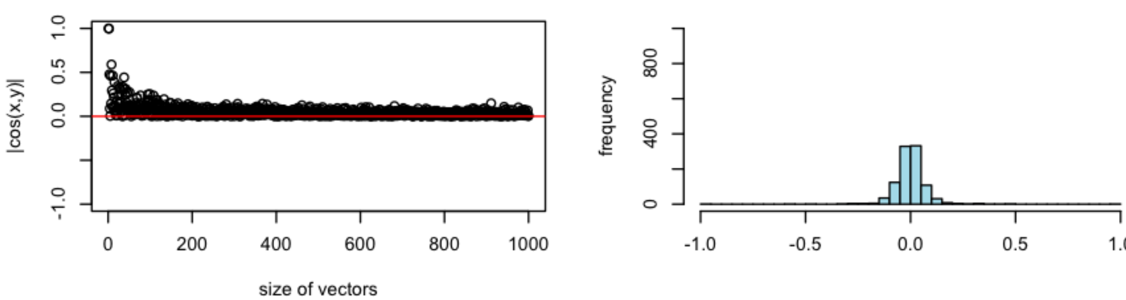

FIGURE 18 – The size of the vectors goes from 1 to 1000

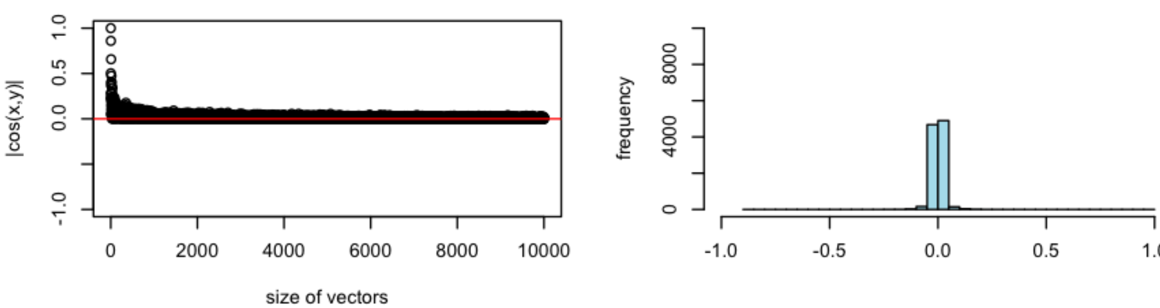

FIGURE 19 – The size of the vectors goes from 1 to 10000

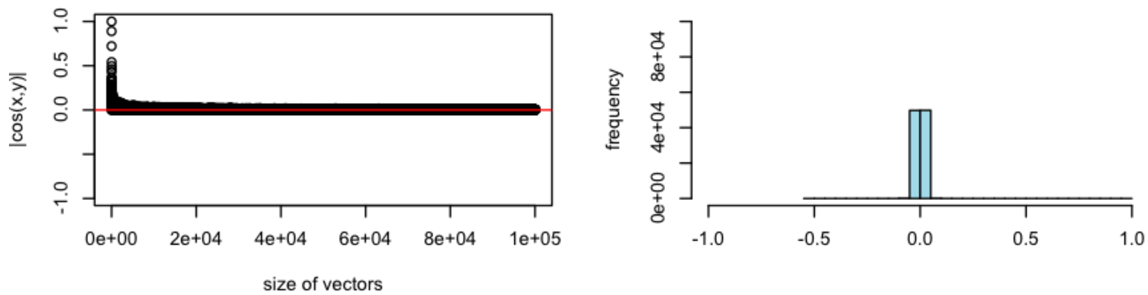

FIGURE 20 – The size of the vectors goes from 1 to 100000

N	$c(\ell_2)$	$c(\ell_3)$	$c(\ell_4)$	$c(\ell_5)$	$c(\ell_6)$	$c(\ell_7)$	$c(\ell_8)$	$c(\ell_9)$	$c(\ell_{10})$
10	0.3914833	0.7593072	0.04978319	0.4778023	0.342336	0.1080792	0.6815922	0.228317	0.4305548
100	0.1111445	0.2198388	0.1944975	0.008304239	0.2429919	0.1836851	0.09046331	0.1741768	0.1219234
1000	0.09502381	0.02903221	0.04312589	0.03447621	0.00332147	0.04386698	0.07434009	0.04587727	0.006665223
10000	0.004675528	0.03526224	0.004552958	0.006673089	0.006159439	0.003180619	0.003608415	0.002587822	0.003079349
100000	0.004063001	0.005613358	0.01064448	0.006071409	0.008743094	0.001314857	0.006016169	0.00519082	0.00111364

TABLE 3 – $\mid \cos(\theta_n) \mid$ for selected n in a sample (each line represents a sample of size N)

2.2.2 Uniform distribution : second case

Suppose that the components of vectors x and y of length n are randomly drawn with uniform distribution law on the interval $[a = -1.0, b = +1.0]$. Let X_j and Y_j represent these random variables corresponding to the sample j and Z_j represents $\cos(\theta_n)$ corresponding to the same sample j (for fixed n). Here, as in the section 2.1.2, we present the mean value $\mathcal{E}_p = \frac{1}{p} \sum_{j=1}^{p} \mid Z_j \mid$ where p is the size of the sample. The left side of the figures 21-25 (resp. 26-30) represents the evolution of $\mathcal{E}_p$ when n is fixed to 10 (resp. 1000) and the size of sample goes from 1 to p (with p = 10, 100, 1000, 10000 and 100000). The right side of the figures 21-25 (resp. 26-30) represents the density of "random variables" $\mathcal{E}_i$, $i = 1, p$. Again, the distribution seems to follow the normal law.

This table 4 shows the decay of $\mathcal{E}_p$ as a function of the growth of p. Each line of the this table presents the mean value of $\mathcal{E}_p$ for $p \in \{10, 100, 1000, 10000, 100000\}$ and a fixed n (10 for the first line and 1000 for the second line). We can notice that $\mathcal{E}_p$ tends to zero when p becomes large.

A zoom (1000 times) on the left side of the last experiment (i.e. : Figure 30) highlights the convergence to zero of $\mathcal{E}_p$. This zoom is presented in the figure 31.

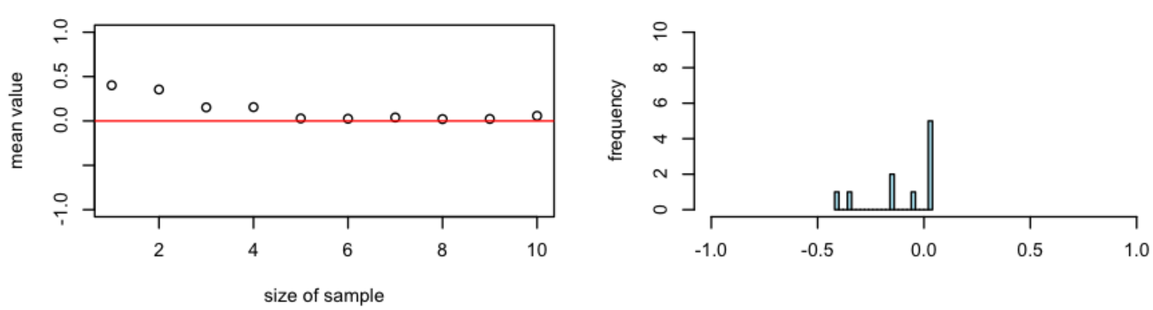

FIGURE 21 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10$ (fixed $n = 10$)

3 Observation

The OR conjecture seems to be verified by the above experiment when the variables are randomly drawn with standard normal and uniform distributions.

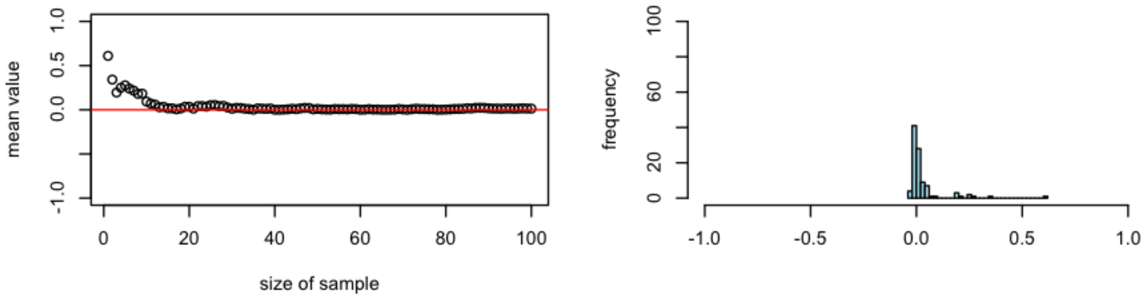

FIGURE 22 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100$ (fixed $n = 10$)

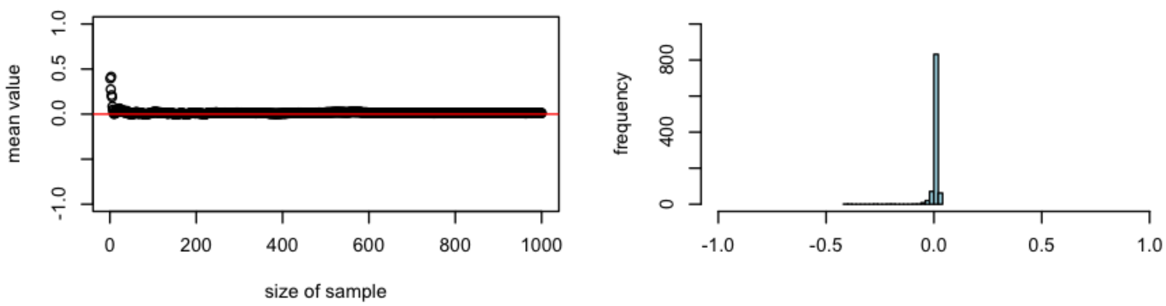

FIGURE 23 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 1000$ (fixed $n = 10$)

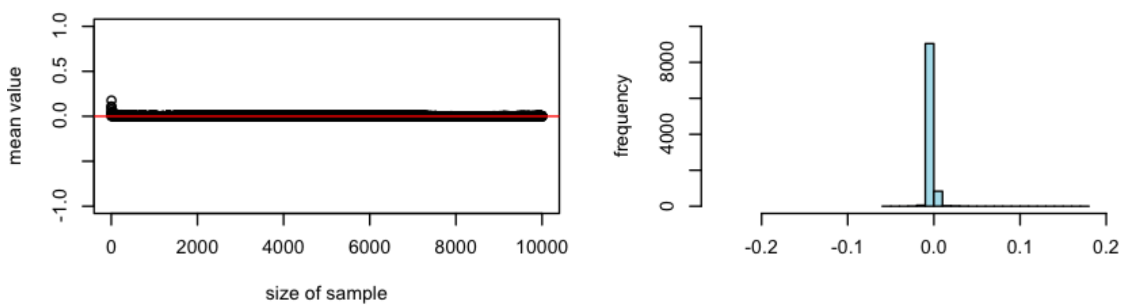

FIGURE 24 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10000$ (fixed $n = 10$)

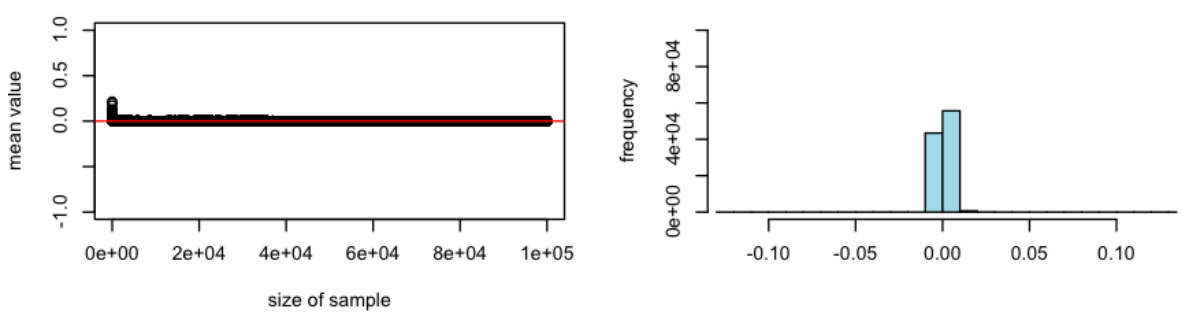

FIGURE 25 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100000$ (fixed $n = 10$)

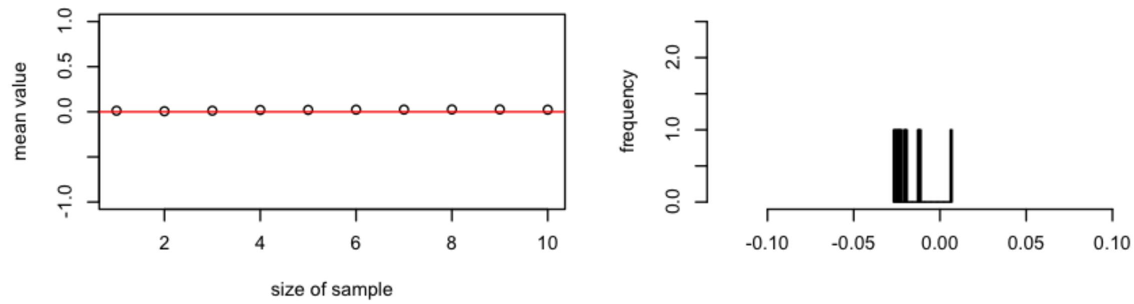

FIGURE 26 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10$ (fixed $n = 1000$)

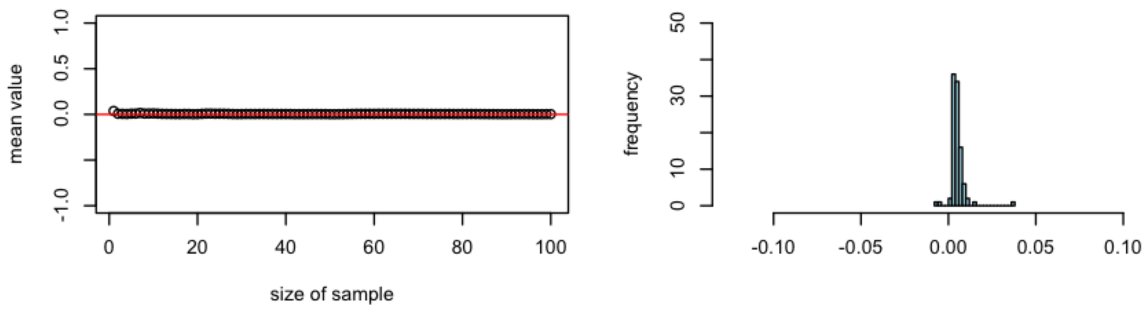

FIGURE 27 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100$ (fixed $n = 1000$)

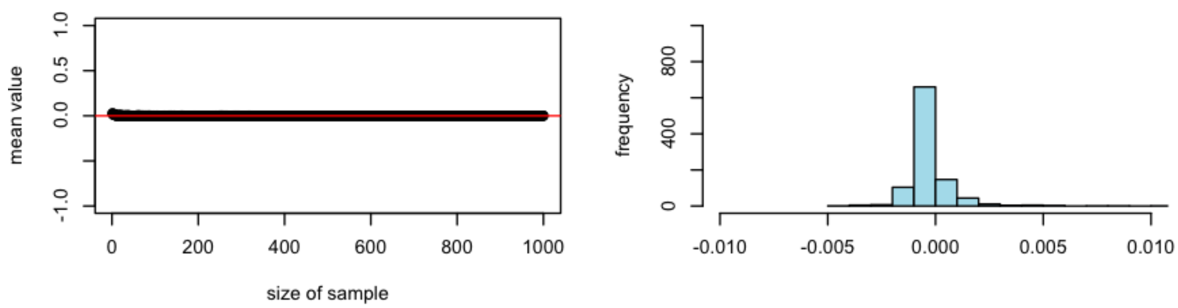

FIGURE 28 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 1000$ (fixed $n = 1000$)

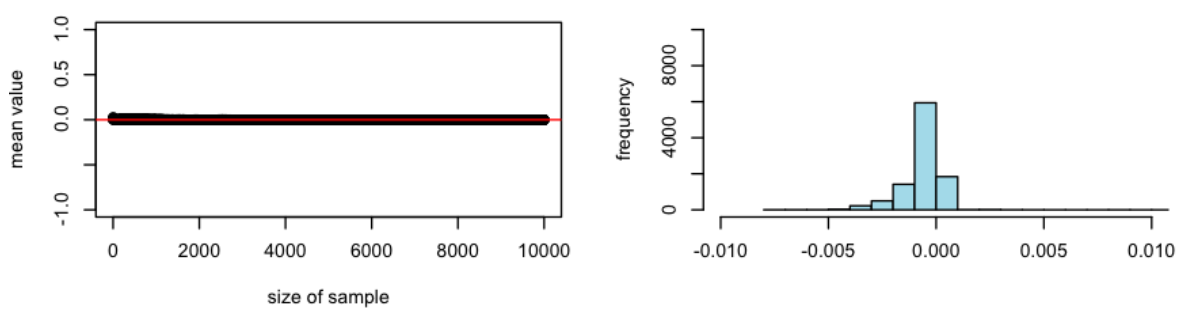

FIGURE 29 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 10000$ (fixed $n = 1000$)

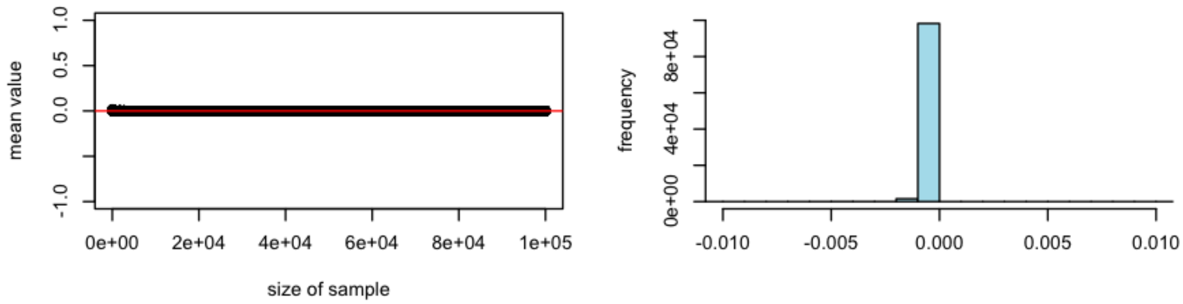

FIGURE 30 – Evolution of the mean value $\mathcal{E}_p$ for $p = 1, 100000$ (fixed $n = 1000$)

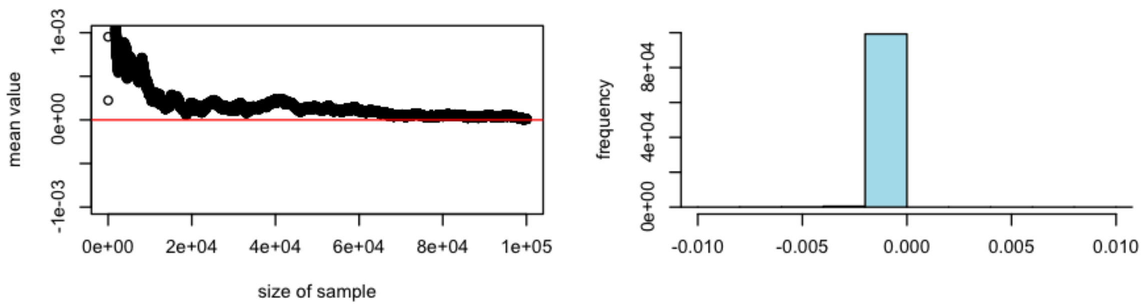

FIGURE 31 – Zoom on the left side of the figure 30

Références

[1] G. Oppenheim and Ph. Ricoux, *Oppenheim-Ricoux (OR) conjecture" Mathias*, 2014.

[2] Maria L. Ritzzo, *Statistical Computing with R"* Chapman & Hall/CRC, 2 edition (July 15, 2015).

N \ p	10	100	1000	10000	100000
10	0.05766342	0.01327352	0.01185789	0.002234513	0.0002172728
1000	0.02290564	0.002709578	0.0007072537	0.0001488402	1.019963e-05

TABLE 4 – $\mathcal{E}_p$ for various n and p