The objectives of this labwork are to study some common random variables and evaluate the quality of their respective estimators using the Monte Carlo simulation.

1. Independence and correlation

In this first part, we study two random variables (RV) X_1 and X_2. In order to characterize them, we compute their probability density function (PDF), their mean and their variance.

We are given two realisations of the random values. Each realisation contains N = 1000 elements. In order to have an estimation of the PDF, we compute the histogram of each realisation. The histogram is indeed an unbiased estimator of the discretized PDF. Here, we generate the histogram with $N_B = 30$ boxes and normalize it so that the integral of the PDF from -∞ to +∞ is 1. If we sum each box of the histogram, we obtain a long box whose area is equal to N×dx, where dx is the size of the support of the box. N×dx is the normalisation factor. We obtain then the estimator of the PDF:

$$\widehat{\mathrm{PDF}} = \frac{\text{histogram of } \mathrm{X_i}}{\mathrm{N\ dx}}$$

where:

$$\mathrm{dx} = \frac{\max(\mathrm{X_i}) - \min(\mathrm{X_i})}{\mathrm{N_B}}$$

The histograms are given on Figures 1 and 2. We compute the mean m and the variance σ of each realisation:

	X_1	X_2
m_i	2,9036	3,2308
σ_i^2	2,0075	2,2419
dx_i	0,286	0,284

For each sample, the mean and the variance are not equal. Given the two histograms, we can consider that X_1 and X_2 are Gaussians. On each histogram, we plot the Gaussian curve with the parameters we found (m_i, σ_i) for each variable (see Figures 1 and 2).

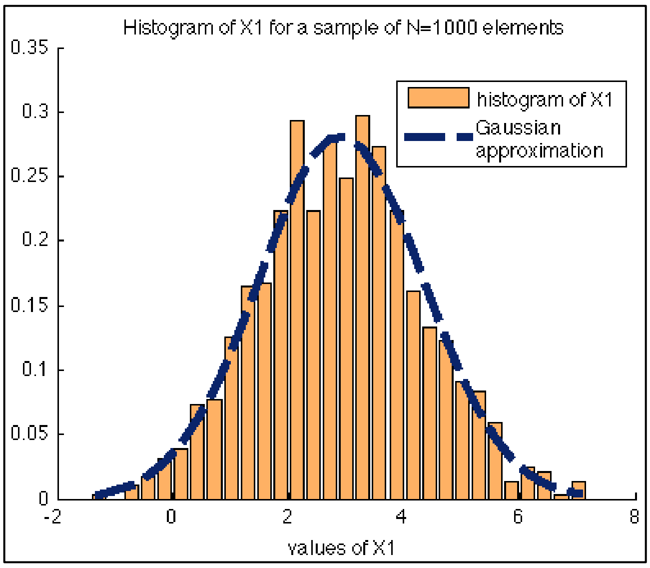

Figure 1: Estimation of the PDF of X1 (histogram of a sample of 1000 elements)

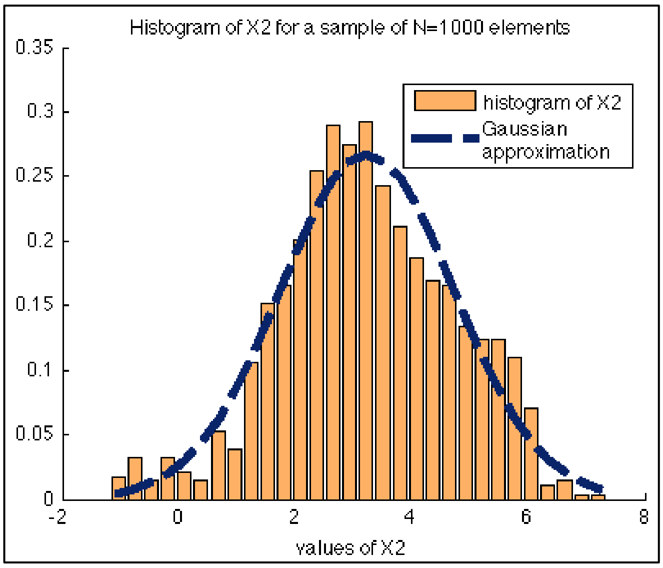

Figure 2: Estimation of the PDF of X2 (histogram of a sample of 1000 elements)

To know whether the elements are independent or not i.e. whether the noise is white or not, we compute the autocorrelation function (ACF) for each RV (see Figures 3 et 4).

Each ACF has a peak. The peak width corresponds to the correlation length L_C. If L_C is equal to one sample size, then the noise is white. It means that it is an independent process.

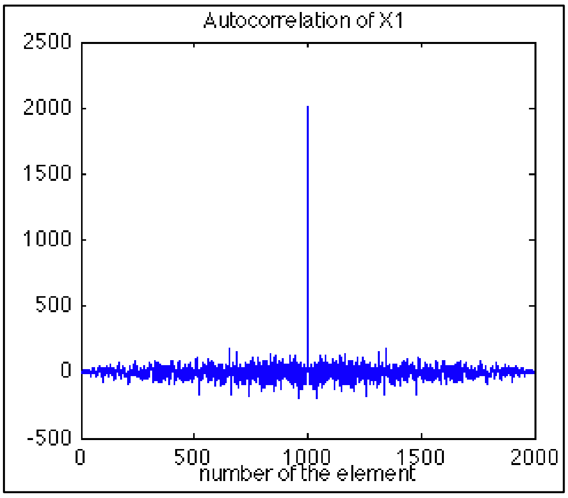

Figure 3: Autocorrelation function of X_1

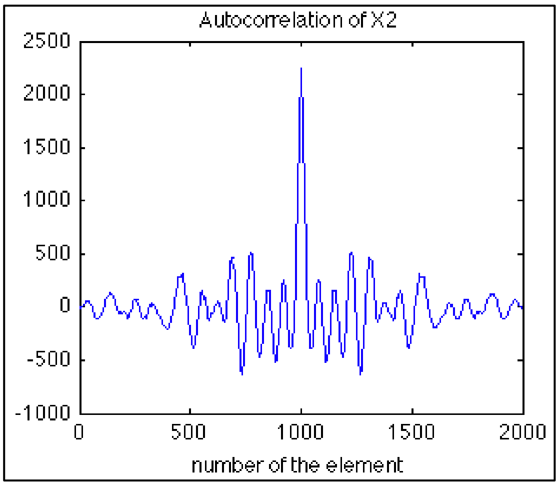

Figure 4: Autocorrelation function of X_2

Here the correlation lengths are:

	X_1	X_2
$L_{c,i}$	1 sample	50 samples

We can consider that the set of samples of X_1 are independent whereas the samples of X_2 are not independent even the correlation length is not high (under 100 samples).

2. Determination of the nature of a random noise

In this part, the signal considered is a noisy image (see Figure 5). In this case, the random process is not stationary. There are indeed three distinct parts whose characteristics are different (PDF, σ, m, L_c…).

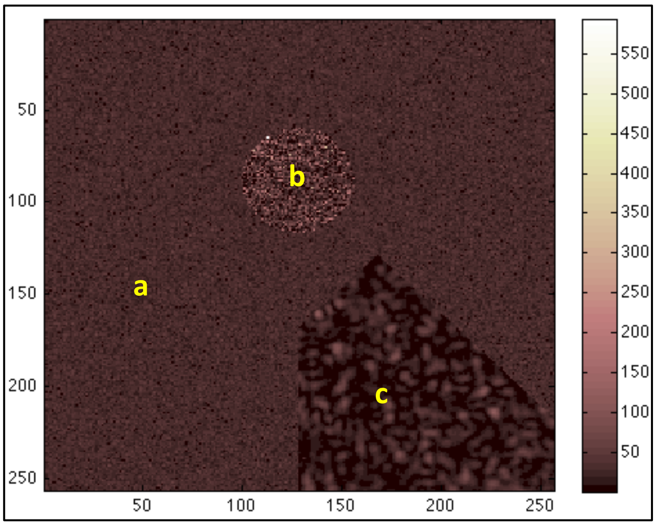

Figure 5: Noisy image with 3 parts of different noise (a, b and c)

We compute the histogram of each noise (see Figure 6, 7 and 8). We have then an estimation of each probability density function that gives us the nature of the noise:

	Noise a	Noise b	Noise c
Nature	Gaussian	Exponential	Gamma
m	30,01	49,15	17,04
σ	10,03	48,74	13,99
white	yes	yes	no

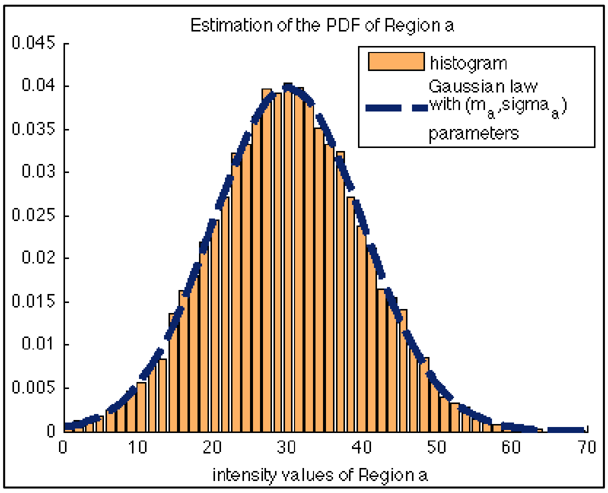

Figure 6: Estimation of the PDF of region b (for N_a=19588 points)

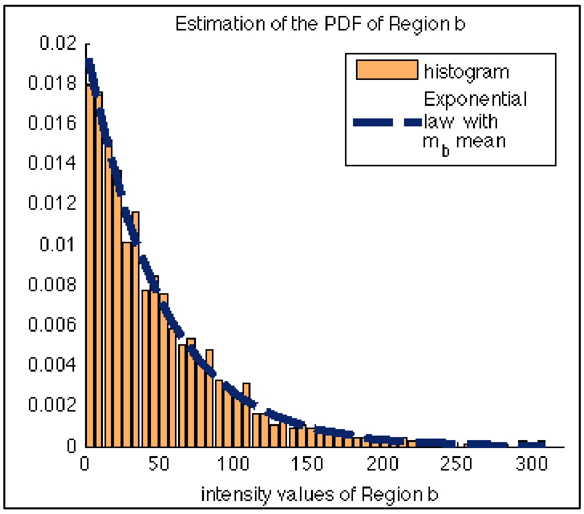

Figure 7: Estimation of the PDF of region b (for N_b=1188 points)

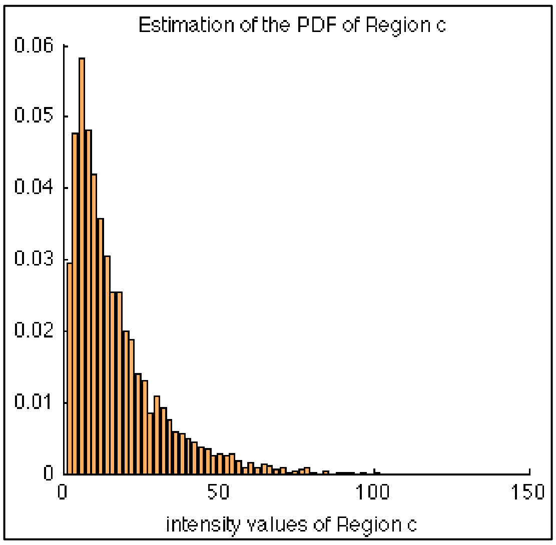

Figure 8: Estimation of the PDF of region b (for N_c=4819 points)

<u>Note:</u> The exponential law is a Gamma law of order 1. It usually corresponds to a speckle intensity PDF. If we sum n independent speckles, we obtain a n-order Gamma law, as Case 3 is not simple to estimate.

3. Variance of an estimator

Estimator of the mean of a Gaussian sample:

We first generate a sample of N = 100 Gaussian elements g with a m=3 mean and a σ=2 standard deviation. The empirical mean is the estimator defined as:

$$\widehat{m} = \frac{1}{N}\sum_{i=1}^{N} g_i$$

Theoretically, the mean of the estimator is:

$$E[\widehat{m}] = m = 3$$

so that this estimator is accordingly unbiased. Its variance is

$$\mathrm{Var}[\widehat{m}] = \frac{\sigma^2}{N} = 0{,}040$$

We test the performances of the estimator with the Monte-Carlo simulation. We generate K realisations of N elements. Then the empirical mean of each realisation is computed: we obtain a set K values. We compute then the mean m_{MC} and the variance var_{MC} of this set.

For N = 100 and K = 1000, we have:

$$\mathbf{m_{MC} = 3{,}103}$$

$$\mathbf{var_{MC} = 0{,}039}$$

We see that if N increases, m_{MC} converges toward m and the variance var_{MC} decreases in 1/N (see Figures 9 and 10).

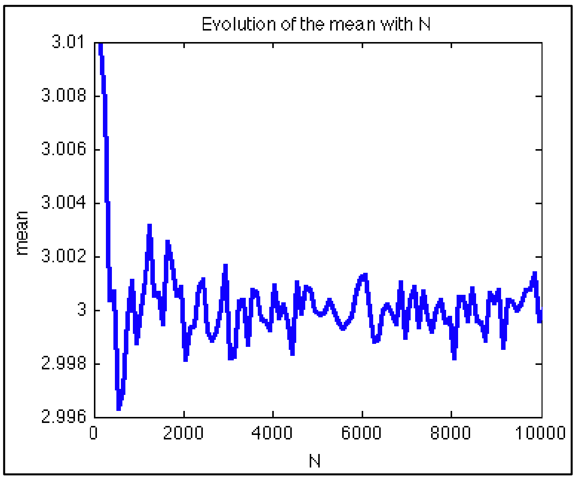

Figure 9 : Evolution of the mean of the estimator with N, with the Monte-Carlo simulation for K=1000 realisations

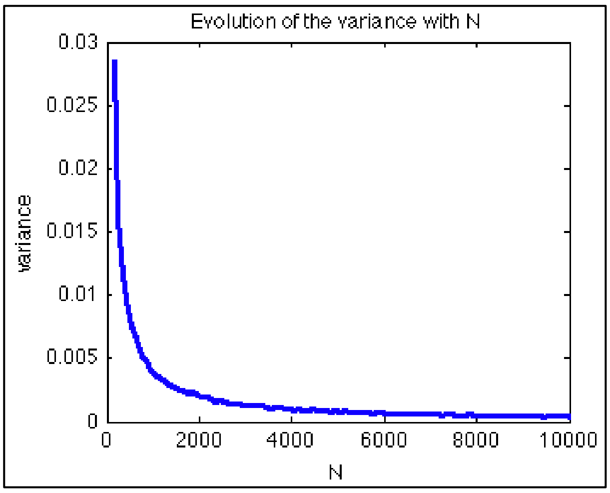

Figure 10 : Evolution of the variance of the estimator with N, with the Monte-Carlo simulation for K=1000 realisations

Moreover, if K tends to infinity, we verify that:

- The mean m_{MC} seems to converge toward m : we see it by computing the variations of the bias with K,
- The variance var_{MC} converges toward $\frac{\sigma^2}{N}$,

(as shown in Figures 11 and 12).

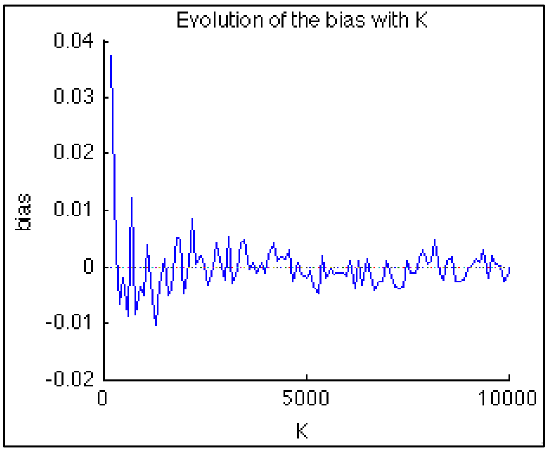

Figure 11 : Variations of the bias with K (N = 100) with the Monte-Carlo simulation

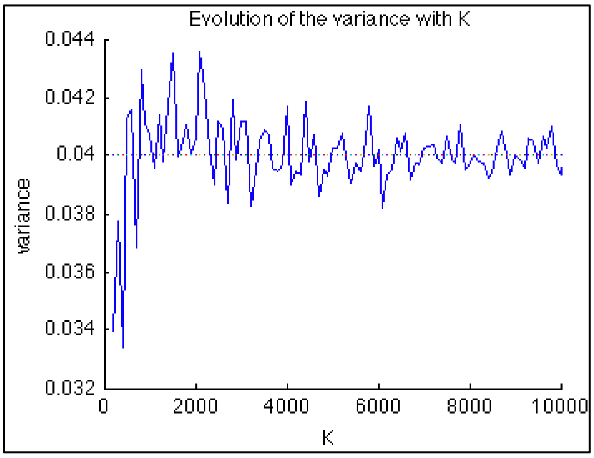

Figure 12: Variations of the variance with K (N = 100) with the Monte-Carlo simulation

Estimator of the mean of an exponential sample:

We generate now a sample of N = 100 independent exponential elements with mean m and standard deviation σ = m. We use the same estimator as for the Gaussian random variable:

$$\hat{m} = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i$$

The estimator bias is zero and its variance is

$$\mathrm{Var}[\hat{m}] = \frac{\sigma^2}{N} = \frac{m^2}{N}$$

We test the performances of the estimator with the Monte-Carlo simulation depending on the value of m (see Figure 13 for the exponential sample and Figure 14 for the Gaussian sample)

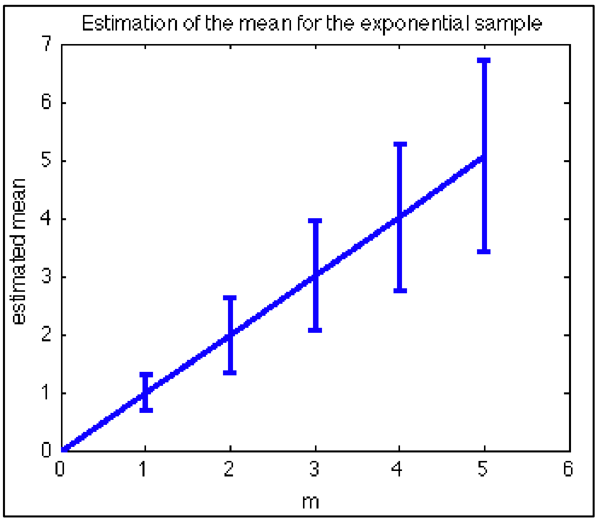

Figure 13: Estimation of the mean of the estimator with its error bar depending on the value of m (N = 100 and K = 1000) Exponential sample

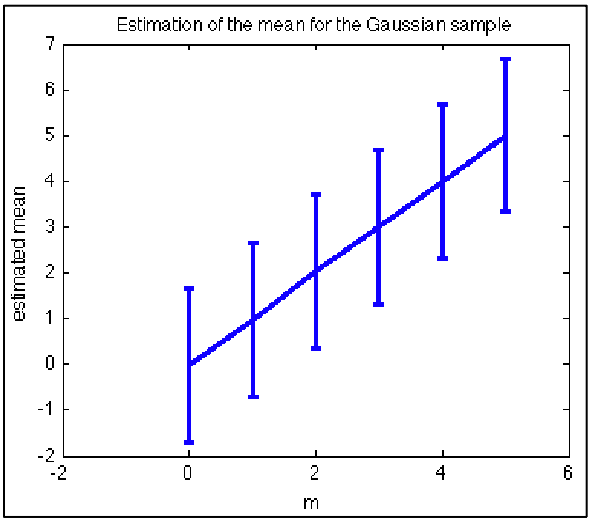

Figure 14: Estimation of the mean with its error bar depending on the value of m (N = 100 and K = 1000) Gaussian sample

The variance of the Gaussian is constant because it does not depend on m. The error bar of the exponential RV increase with m since σ = m.

4. Estimation of the location parameter of a Cauchy RV

Let us consider now a Cauchy random variable whose probability density function is:

$$P_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - a)^2}$$

The goal is to estimate the parameter a.

Since the function $x \cdot P_X(x)$ is not absolutely integrable[1], the mean, the variance or the moments are not computable. If we try to compute the empirical mean, we will see that it will not converge toward any value. Therefore it is not a reliable estimator.

However, we can still estimate the parameter a by using the median estimator, since its median is defined and equal to a. We can verify it by calculating the following integral:

$$\int_{-\infty}^{a} \frac{1}{\pi} \cdot \frac{1}{1 + (x - a)^2} dx = \frac{1}{\pi} \cdot \int_{-\infty}^{0} \frac{dt}{1 + t^2} = \frac{1}{2}$$

Consequently this estimator is unbiased.

In order to illustrate it, we generated K=1000 samples of N elements. Th Monte-Carlo simulation gives us the median with its standard deviation. We plot the estimation of the mean too, using the estimator $\widehat{m}$ (see Figure 15). Figure 16 shows that the estimator of the median converges rapidly toward 'a' (or the median). It is also consistent to estimate the parameter a.

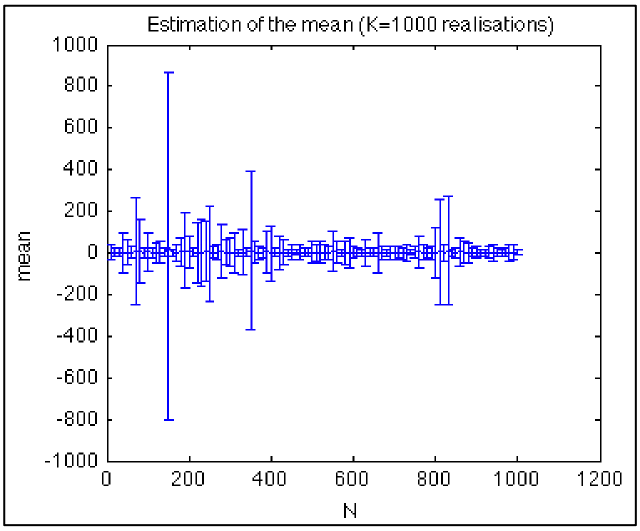

Figure 15 : Variations of the estimation of the mean with N (K = 1000 realisations)

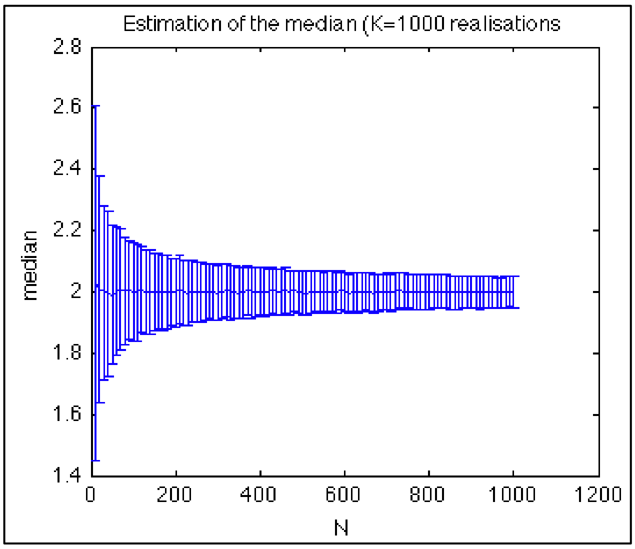

Figure 16: Variations of the estimation of the median with N (K = 1000 realisations)

Conclusion

To characterize a random variable or a noise, it is useful to compute the PDF and the ACF : the PDF gives information about the nature of the noise (the probability law, the mean, the variance…) and The ACF gives information about the stationarity (if the noise is white or not).

This labwork was also the occasion for us to manipulate the empirical mean estimator with gaussian, exponential and Cauchy random variables. In order to test estimators, we use the Monte-Carlo simulation which is, in these cases, easy to implement. It gives the parameters of the estiamtor (mean, variance…) if the numbers of samples is high. With this simulation, we could see that sometimes an estimator could not work with some RV (mean estimator for Cauchy RV…) and we need to use another one (the median estimator).

The quality of an estimator is its capability to give estimate the parametr we want, the way it converges to it, its standard devaition and its compatibility with the noise: we choose the estimator regarding the bias and variance

[1] The function $x \cdot P_X(x)$ is equivalent to $\frac{1}{x}$ as x tends do infinity. Consequently, the integral $\int_{-\infty}^{\infty} |x\, P_X(x)| dx$ does not exist.