Facultat de Física de la Universitat de Barcelona
Probability and Statistics

Task 2

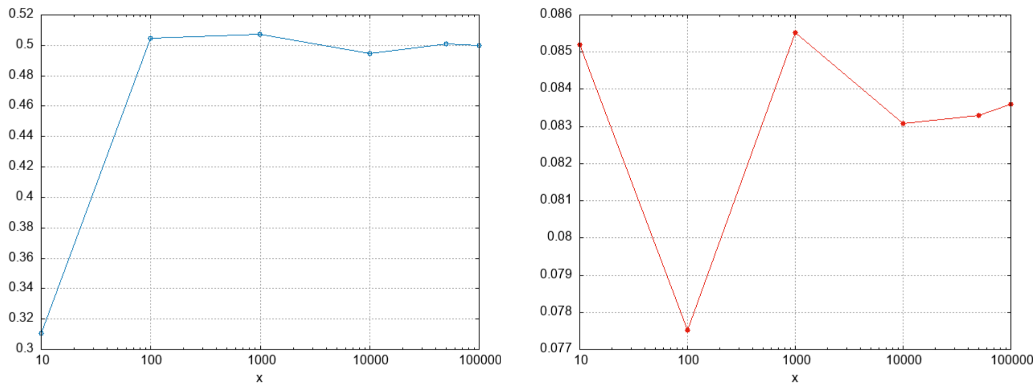
Javier Castillo Uviña

jcastiuv7@alumnes.ub.edu

1.

It is computed the $S_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ and variance objects for samples of $n = 10$, 100, 1000, 10000, 50000, 100000 random numbers for the following distributions:

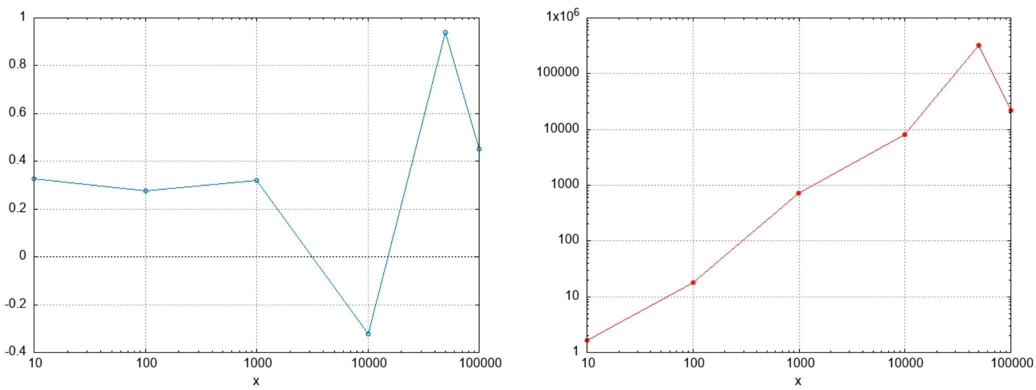
1. Uniform distribution in the interval $[0, 1]$:



1 – Mean **2 – Variance**

Figure 1 – Uniform parameters convergence

2. Cauchy distribution $p(x) = \frac{1}{\pi(1+x)^2}$:



1 – Mean **2 – Variance**

Figure 2 – Cauchy parameters convergence

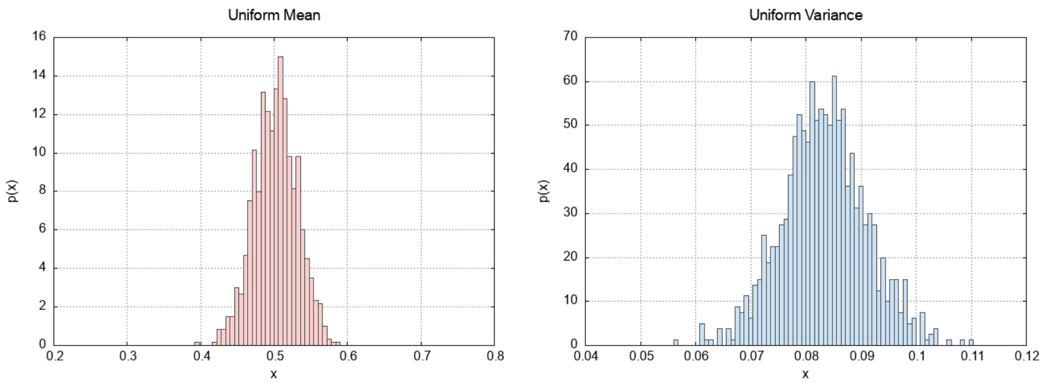
3. As we can see the bigger the data set the accuracy the parameter value for the well defined parameters distribution. In Figure 11 we can see the mean value of a $U[0, 1]$ is

0.5. In Figure 12 it also converges to the analytic value $\frac{1}{12}$. On the other hand neither the Cauchy mean nor the Cauchy Variance seems to converge to a specific value, what's more, in Figure 22 we can see the Cauchy variance strongly diverges.

2.

For the following distributions it is generated $m = 1000$ data sets of $n = 100$ random values each. After computing the mean and variance of each data set they will be sampled in order to see their distribution.

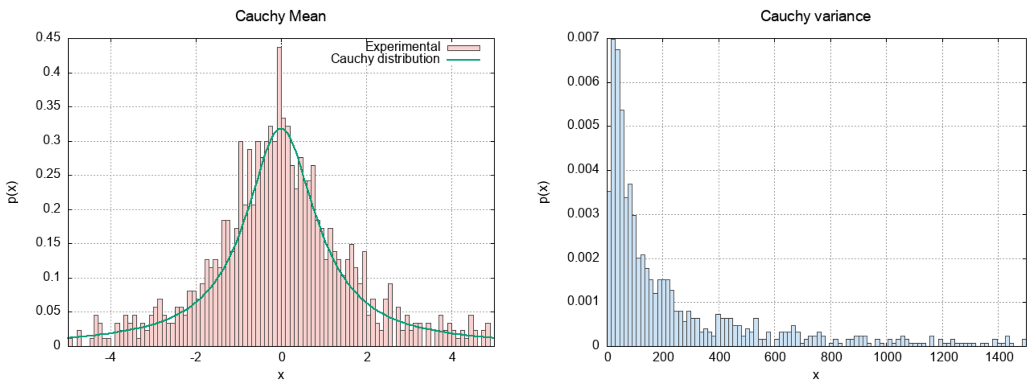
1. Uniform distribution in the interval $[0, 1]$:



1 – Mean **2 – Variance**

Figure 3 – Uniform parameters distribution

2. Cauchy distribution $p(x) = \frac{1}{\pi(1+x)^2}$:



1 – Mean **2 – Variance**

Figure 4 – Cauchy parameters distribution

3. In Figure 3 is observable how the each uniform parameter distribution is a Gaussian with μ equal to the analytic value. For the Cauchy case (Figure 4) only the mean parameter value seems converge. The Cauchy mean distribution is again a Cauchy distribution with the same parameters than the primitive one, as we can see in 41. This is due to the fact that the Cauchy distribution doesn't have the $n - order$ moments well defined. It also implies that the Cauchy variance distribution doesn't obeys the CLT.

3.

Being $P_{X_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} exp\Big(- \frac{(x-\mu_1)^2}{2\sigma_1^2}\Big)$ and $P_{X_2}(x) = \frac{1}{\sqrt{2\pi}\sigma_2} exp\Big(- \frac{(x-\mu_2)^2}{2\sigma_2^2}\Big)$, two gaussian distributions for two different random variables lets compute $P_{X_1}(x) * P_{X_2}(x)$:

$$P_{X_1}(x) * P_{X_2}(x) = P_{X_1+X_2}(x) = \int_{-\infty}^{\infty} P_{X_1}(x-x')P_{X_2}(x')dx' = \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{\varphi(x,x')}dx'$$

Were $\varphi(x,x') = -\frac{(x-x'-\mu_1)^2}{2\sigma_1^2} - \frac{(x'-\mu_2)^2}{2\sigma_2^2}$.

It is proposed the variable change:

$$\begin{aligned}\varphi(x,x') &= -\frac{(x-x'-\mu_1)^2}{2\sigma_1^2} - \frac{(x'-\mu_2)^2}{2\sigma_2^2} \\ &= -\frac{(x-\mu_1)^2\sigma_2^2 + (\sigma_1^2+\sigma_2^2)x'^2 - 2x'[(x-\mu_1)\sigma_2^2 + \mu_2\sigma_1^2] + \mu_2^2\sigma_1^2}{2\sigma_1^2\sigma_2^2} \\ &= \Big\{\sigma = \sqrt{\sigma_1^2+\sigma_2^2}\Big\} = -\frac{x'^2 - 2x'\frac{(x-\mu_1)\sigma_2^2+\mu_2\sigma_1^2}{\sigma^2} + \frac{(x-\mu_1)^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma^2}}{2\frac{\sigma_1^2\sigma_2^2}{\sigma^2}}\end{aligned}$$

So we can separate this expression in two exponential:

$$\frac{1}{\sqrt{2\pi}\sigma} exp\Bigg[- \frac{(x-\mu_1)^2\sigma_2^2 + \mu_2^2\sigma_1^2}{2\sigma_1^2\sigma_2^2} - \frac{[(x-\mu_1)\sigma_2^2 + \mu_2\sigma_1^2]^2}{2\sigma_1^2\sigma_2^2\sigma^2}\Bigg]\cdot$$
$$\cdot\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{\sigma_1^2\sigma_2^2}{\sigma}} exp\Bigg[- \frac{\Big(x' - \frac{(x-\mu_1)\sigma_2^2+\mu_2\sigma_1^2}{\sigma^2}\Big)^2}{2\frac{\sigma_1^2\sigma_2^2}{\sigma^2}}\Bigg]dx'$$

The valu of this integral is equal to 1 so it is proved that:

$$P_{X_1}(x) * P_{X_2}(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\Bigg[- \frac{(x-\mu_1-\mu_2)^2}{2\sigma^2}\Bigg]$$

Which is the expression for a gaussian with $\mu = \mu_1 + \mu_2$ and $\sigma = \sqrt{\sigma_1^2+\sigma_2^2}$.

4.

In order to demonstrate that that the Central Limit Theorem implies the Weak Law of Large Numbers. We define $S_N = \sum_{k=1}^{n} X_k$, then notice $\mu_{S_n} = n\mu$, Var[S_n]= $n\sigma^2$ so $\sigma_{S_n} = \sigma\sqrt{n}$. Let's suppose then:

$$\begin{aligned}\left|\frac{S_n}{n} - \mu\right| > \epsilon &\Rightarrow |S_n - n\mu| > n\epsilon \\ &\Rightarrow \left|\frac{S_n - n\mu}{\sigma_{S_n}}\right| > \frac{n\epsilon}{\sigma_{S_n}} \\ &\Rightarrow |S_n^*| > \frac{\sqrt{n}\epsilon}{\sigma} > x\end{aligned}$$

we consider x any real positive number for n large enough. So now we have:

$$P\left\{\left|\frac{S_n}{n}-\mu\right|>\epsilon\right\}=P\left\{|S_n^*|>\frac{\sqrt{n}\epsilon}{\sigma}\right\}\leq P\{|S_n^*|>x\}$$

So

$$lim_{n\to\infty}P\{|S_n^*|>x\}=lim_{n\to\infty}P\{S_n^*>x\}+lim_{n\to\infty}P\{S_n^*<-x\}=1-F(x)+F(-x)$$

This $F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}dy$ comes from the Central Limit Theorem. CLT states: Let $S_N=\sum_{k=1}^{n}X_k$ where $X_1,X_2\cdots,X_n$ are mutually independent variables with the same mean, μ, and deviation, σ, and let S_n^* be the normalized version of S_n. Then

$$P\{S_n^*\leq x\}=F(x)$$

So if we choose an x large enough we find $F(x)=1$ and $F(-x)=0$ that's how we finally find

$$lim_{n\to\infty}P\left\{\left|\frac{S_n}{n}-\mu\right|>\epsilon\right\}\leq lim_{n\to\infty}P\{|S_n^*|>x\}=0$$

As it is shown below the Central Limit Theorem implies the Weak Law of Large Numbers.

5.

It is generated $m=1000$ sets of $n=100$ random values distributed uniformly in the interval $[0,1]$. It is found the maximum value of each set and it is plotted in 5. This analytic behaviour is found by:

$$P(Z\leq z)=P(X_1\leq z,X_2\leq z,\ldots,X_n\leq z)=P(X_1\leq z)P(X_2\leq z)\cdots P(X_n\leq z)=$$
$$=\prod_{i=1}^{n}F_X(z)=(F_X(z))^n=z^n$$

where $Z=max(X_1,X_2,\ldots,X_n)$.

So it is obtained:

$$p(z)=nz^{n-1}$$

This $p(z)$ is what we call analytic behaviour in the graphics below. The limit EVD is a Weibull with $\alpha=1$. so the final expression is:

$$p(x)=ne^{-n(1-x)}\ 0\leq x\leq 1$$

The probability of having a maximum greater than 0.9999 for this case is:

$$P(Z>0.9999)=1-P(Z\leq 0.9999)=1-0.9999^n=0.00995\mid n=100$$

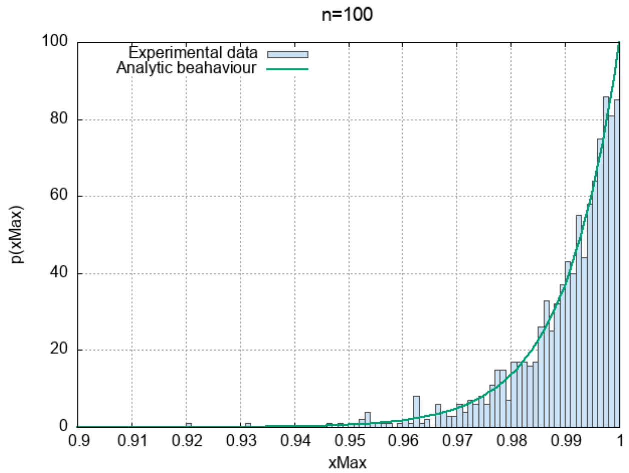


Figure 5 – Extreme value distribution for $m = 1000$ maximums for sets of $n = 100$ random values

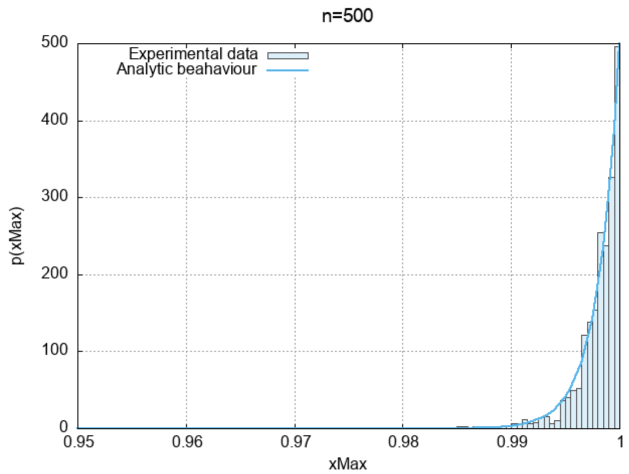


Figure 6 – Extreme value distribution for $m = 1000$ maximums for sets of $n = 500$ random values

The probability of having a maximum greater than 0.9999 in the n=500 case is:

$$P(Z > 0.9999) = 1 - P(Z \leq 0.9999) = 1 - 0.9999^n = 0.04877 \mid n = 500$$