Strong Law of large Numbers and ...

Theorem (Strong Law of Large Number ($SLLN$))

▷ *Let X_k, $k \geqslant 1$, be a sequence of i.i.d. copies of X such that $\mathbb{E}\,X$ is well-defined Then*

$$\overline{X}_M := \frac{1}{M}\sum_{k=1}^{M} X_k \xrightarrow{a.s.\&L^1} m_X = \mathbb{E}\,X \quad \text{as} \quad M \to +\infty.$$

▷ *More genreally, if $\mathbb{E}\,\psi(X)$ is well-defined, then*

$$\frac{1}{M}\sum_{k=1}^{M} \psi(X_k) \xrightarrow{a.s.\&L^1} \mathbb{E}\,\psi(X) \quad \text{as} \quad M \to +\infty.$$

- Notations: $:=$ stands for a the definition of an object.
- This suggests to go back on the scenario space ω and use simulation to approximate $\mathbb{E}\,X$.
- Physical device...? Not realistic...
- On a computer ? Needs normalization.

- How to work on the scenario space, even if X takes infinitely many values ?
- Write $X = \varphi(U)$, U uniformly distibuted over $[0,1]$ or $[0,1]^d$ where φ is made explicit, X can be simulated and ...
- Note that if $U = (U^1, \dots, U^d)$ has $U([0,1]^d)$-distribution $\iff U^1, \dots, U^d$ are independent and $U([0,1])$-distributed.
- Why not use (repeated) simulation?
- Hence everything relies on simulating independent realizations of the simplest continuous law:

$$U([0,1]) = \mathbf{1}_{[0,1]}(\xi)d\xi)$$

- Once again, how?
- History: the idea goes back to Buffon (Buffon's needle) and for practical implementation to Stanislas Ulam during WW II for atomic and then H-bomb.

Simulation: random (?) numbers

- What should they be: not so clear. . . Say, binary digits
- Physical device to "draw"/"sample" at random sequences of 0 and 1?
- Not reliable and, in any case, too slow for computational matter.

Switch to random numbers generated on a computer?

- Little theoretical suport.
- Random $\Rightarrow$ algorithm complexity $= O(\text{sample size})$.
- Logic approach: any specified sequence of numbers is not random.

Leads to/suggests a less ambitious statistical approach . . .

Pseudo-random numbers I

- A sequence $(x_n)_{1\leqslant n\leqslant N}$ of pseudo-random numbers is
  - an as perfect as possible imitation of a generic sample path of $(U_n(\omega))_{n\geqslant 1}$, U_n i.i.d. with $\mathcal{U}([0,1])$-distribution (defined on $(\Omega,\mathcal{A},\mathbb{P})$),
  - generated on a computer
  - with a low computational cost.

- What does generic mean ? The scenario ω lies in any event of probability 1 of interest...
  - Not so clear because $N < +\infty$!!
  - whereas asymptotic statistical frequentist properties are requested (averaging, de-correlation, etc)
  - In particular, for every not too large d,

$$\big(x_{d(n-1)+1},\ldots,x_{d(n-1)+d}\big)_{n\leqslant N/d} \text{ is a good imitation of i.i.d.copies of } \mathcal{U}([0,1]^d)$$

  - Toolbox DIEHARD (Marsaglia, 1998).

Pseudo-random numbers II

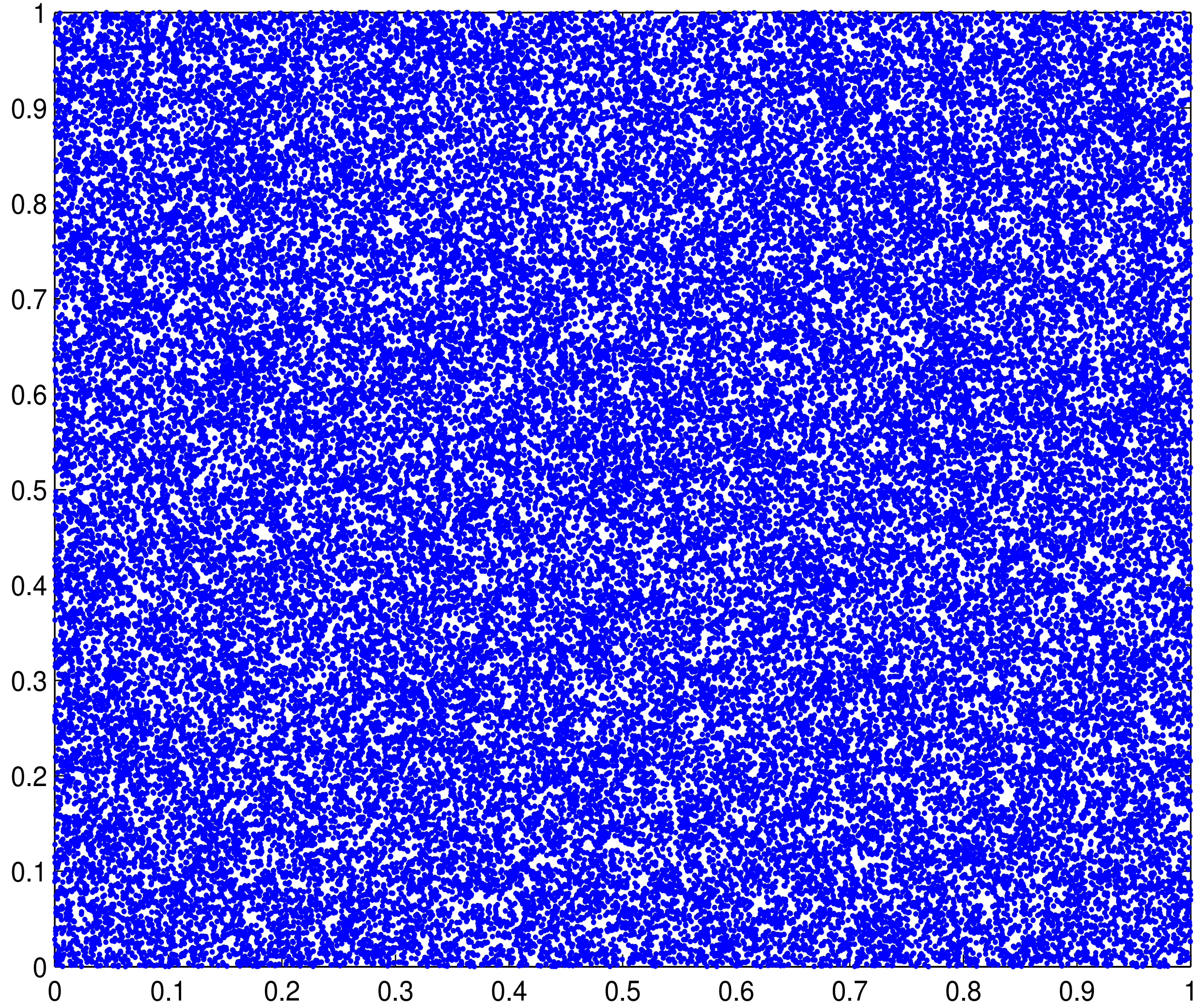
- By arithmetic means, one may compute congruational pseudo-random numbers of the form
  - $x_n = \frac{y_n}{N}, \quad y_{n+1} = a\, y_n + b \mod N,\ a \not\equiv 0 \mod N,$
  - N prime $\Rightarrow o(a) = N - 1$ in $(\mathbb{Z}/N\mathbb{Z})^*$.
- Needs good choices of (very large) prime N, $a \in \{1, \ldots, N-1\}$ and b,
- N is a prime Mersenne numbers (among $2^p - 1$, p prime) *e.g.*

$$N = 2^{19937} - 1.$$

- $\rightsquigarrow$ Mersenne twister random number generators.

  `www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html`
- Standard random numbers generators for FORTRAN, C++, Matlab, Python, etc, have improved.

60 000 pseudo-random numbers (texture)



Fundamental theorem of simulation: theory and practice

- Theory:

Theorem (Fundamental theorem of simulation)

Any random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ Polish space can be written

$$X = \varphi(U), \quad U \stackrel{d}{=} \mathcal{U}([0,1]) \text{ where } \varphi : [0,1] \to (E, \mathcal{E}).$$

- Practice:
  - If $F_X(u) = \mathbb{P}(X \leqslant u)$, $u \in \mathbb{R}$, cumulative distribution function (c.d.f.),

$$X \stackrel{d}{=} (F_X)^{-1}_{left}(U), \qquad U \stackrel{d}{=} \mathcal{U}([0,1]).$$

  - Thus $X \stackrel{d}{=} \mathcal{E}(\lambda)$, $F_X(x) = (1 - e^{-\lambda x})\mathbf{1}_{\{x \geqslant 0\}}$, $\lambda > 0$, so that

$$X \stackrel{d}{=} -\frac{\log(1-U)}{\lambda} \stackrel{d}{=} -\frac{\log(U)}{\lambda}.$$

- Usually, such a φ is not known. More often there exists a computable φ such that

$$X = \varphi(U), \quad U \stackrel{d}{=} \mathcal{U}([0,1]^d) \quad \text{or} \quad \stackrel{d}{=} U\big([0,1]^{(\mathbb{N}^*)}\big).$$

Ex: Box-Muller method for bi-variate normal distribution (see later on)

$$\sqrt{-2\log U_1}\big(cos(2\pi U_2), \sin(2\pi U_2)\big) \stackrel{d}{=} \mathcal{N}\big(0; I_2\big).$$

- $[0,1]^{(\mathbb{N}^*)} = \bigcup_{d\geqslant 1}[0,1]^d$ set of finite $[0,1]$-valued sequences.

$$\varphi(u) = \Psi(\tau(u), u^1, \ldots, u^{\tau(u)}) \text{ where } \tau^{-1}(k) \in \mathcal{B}or([0,1]^k)$$

corresponds to Von Neumann Acceptance-Rejection method.

Ex: Polar method (Marsaglia), compete with Box-Muller for $\mathcal{N}\big(0; I_2\big)$.

- However, usually, no closed form, no approximation by cubature formula (due to high dimension d).

Definition (Monte Carlo simulation “0”)

Monte Carlo simulation can be “defined” as

(1) Representation of X as $X = \varphi(U)$, $U \stackrel{d}{=} \mathcal{U}([0,1]^d)$, $d \in \mathbb{N}^* \cup \{\infty\}$

(2) Generate M independent copies $X_k = \varphi(U_k)$, $(U_k)_{k\geqslant 1}$ i.i.d. with distribution $\mathcal{U}([0,1]^d)$

(3) Compute $d \times M$ independent copies $u_1, \ldots, u_{dM}$ of the uniform distribution over $[0,1]$ using pseudo-random numbers, packed d by d as follows $(u_{(k-1)d+1:kd})$, $k = 1 : M$. Compute the arithmetic mean

$$m_X \approx \frac{1}{M}\sum_{k=1}^{M} \varphi\big(u_{(k-1)d+1:kd}\big)$$

(4) Forget about “pseudo-” in pseudo-random numbers of step (3) and think about $u_k = U_k(\omega)$ and $\overline{X}_M(\omega)$, for a generic ω...

- Question. M how large ? (with $M \ll N/d$ of course! Say $\leqslant 10\% N/d$)

Rates of convergence

- $\overline{X}_M$ as an unbiased estimator of m_X.

$$\mathbb{E}\,\overline{X}_M = \frac{1}{M}\sum_{k=1}^{M}\mathbb{E}\,X_k = m_X.$$

- Quadratic (L^2-rate of convergence: *MSE* and *RMSE*).
  It is a non asymptotic quadratic exact rate due to independence (no correlation is enough)

$$\begin{aligned}
\|\overline{X}_M - m_X\|_2^2 &= \|\overline{X}_M - \mathbb{E}\,\overline{X}_M\|_2^2 = \mathbb{E}\left(\overline{X}_M - \mathbb{E}\,\overline{X}_M\right)^2\\
&= \mathrm{Var}\left(\overline{X}_M\right)\\
&= \frac{1}{M^2}\sum_{k=1}^{M}\mathrm{Var}(X_k) = \frac{1}{M^2}M\,\mathrm{Var}(X)\\
\|\overline{X}_M - m_X\|_2^2 &= \frac{\mathrm{Var}(X)}{M}.
\end{aligned}$$

or, equivalently,

$$\left\|\overline{X}_M - m_X\right\|_2 = \frac{\sigma_X}{\sqrt{M}}$$

with $\underbrace{\sigma_X = \sqrt{\mathrm{Var}(X)}}_{\text{standard deviation}}$ and $\mathrm{Var}(X) = \mathbb{E}\,(X - \mathbb{E}\,X)^2 = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2$.

- Weak rate of convergence.
  The weak convergence is ruled by the Central Limit Theorem (*CLT*)

$$\sqrt{M}\left(\overline{X}_M - m_X\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0;\sigma_X^2) = \sigma_X Z,\ Z \stackrel{d}{=} \mathcal{N}(0;1)$$

  as $M \to +\infty$. We will come back on this crucial result.

- Strong *a.s.* rate of convergence.
  The Law of the Iterated Logarithm (*LIL*) provides an *a.s.* rate of convergence, namely

$$\overline{\lim}_M \sqrt{\frac{M}{2\log(\log M)}}\left(\overline{X}_M - m_X\right) = \pm\sigma_X \quad a.s.$$

Weak convergence (very short background)

- Definition: $(Y_M)_{M\geqslant 1}$ sequence of random variables.

$$Y_M \xrightarrow{\mathcal{L}} Y_\infty \quad \text{as} \quad M \to +\infty$$

if $$\forall f \in \mathcal{C}_{bounded}(\mathbb{R},\mathbb{R}), \quad \mathbb{E}\, f(Y_M) \to \mathbb{E}\, f(Y_\infty)$$

- Extension: Let

$$\mathrm{Disc}(f) = \big\{x \in \mathbb{R} : f \text{ is discontinuous at } x\big\}.$$

Proposition

(a) If $Y_M \xrightarrow{\mathcal{L}} Y_\infty$, $f:\mathbb{R}\to\mathbb{R}$ is bounded and $\mathbb{P}\big(Y_\infty \in \mathrm{Disc}(f)\big) = 0$, then

$$\mathbb{E}\, f(Y_M) \to \mathbb{E}\, f(Y_\infty).$$

(b) Thus, if $f = \mathbf{1}_{[K,K']}$ and $\mathbb{P}(Y_\infty = y) = 0,\ \forall\, y \in \mathbb{R}$, then $\mathrm{Disc}(f) = \{K, K'\}$ so that $\mathbb{P}\big(Y_\infty \in \mathrm{Disc}(f)\big) = 0$, Hence, as $M \to +\infty$,

$$\mathbb{P}(Y_M \in [K,K']) = \mathbb{E}\mathbf{1}_{[K,K']}(Y_M) \longrightarrow \mathbb{E}\,\mathbf{1}_{[K,K']}(Y_\infty) = \mathbb{P}(Y_\infty \in [K,K']).$$

Application to CLT

- How to use the CLT ?

$$\sqrt{M}\,\big(\overline{X}_M - m\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0;\sigma_X^2\big) = \sigma_X Z,\ \ Z \stackrel{d}{=} \mathcal{N}(0;1)$$

- Assume $\sigma = \sigma_X > 0$. For every $z \in \mathbb{R}$,
  $\mathbb{P}(\sigma Z = z) = \mathbb{P}(Z = z/\sigma) = \int_{\{z/\sigma\}} e^{-\frac{\xi^2}{2}} \frac{d\xi}{\sqrt{2\pi}} = 0$. Hence

$$\mathbb{P}\left(\sqrt{M}\,\frac{\overline{X}_M - m_X}{\sigma} \in [b,a]\right) = \mathbb{P}\big(\sqrt{M}\big(\overline{X}_M - m_X\big) \in [\sigma b, \sigma a]\big)$$

$$\xrightarrow{CLT\&(b)} \mathbb{P}(\sigma Z \in [\sigma b, \sigma a]) \quad \text{as} \quad M \to +\infty$$

$$= \mathbb{P}(Z \in [b,a]).$$

Confidence intervals I

- The convergence in distribution of the *CLT* means that for every real numbers $b < a$,

$$\lim_M \mathbb{P}\left(\sqrt{M}\,\frac{\overline{X}_M - m_X}{\sigma_X} \in [b,a]\right) \overset{M\to+\infty}{\longrightarrow} \mathbb{P}(\mathcal{N}(0;1) \in [b,a]) = \Phi_0(a) - \Phi_0(b)$$

where $\Phi_0(x) = \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}}\frac{d\xi}{\sqrt{2\pi}}$ denotes the c.d.f. of $\mathcal{N}(0;1)$.

- Keeping in mind $\Phi_0(x) + \Phi_0(-x) = 1$, $a > 0$, $b = -a$.

$$\lim_M \mathbb{P}\left(\sqrt{M}\,\frac{\overline{X}_M - m_X}{\sigma_X} \in [-a,a]\right) \overset{M\to+\infty}{\longrightarrow} 2\Phi_0(a) - 1 = \mathbb{P}\big(|\mathcal{N}(0;1)| \leqslant a\big).$$

- Confidence level α. Let q_α be the two-sided quantile

$$\mathbb{P}(|\mathcal{N}(0;1)| \leqslant q_\alpha) = \alpha \qquad i.e. \qquad 2\Phi_0(q_\alpha) - 1 = \alpha.$$

- Define the random interval

$$J_{\alpha,M}(\omega) := \left[\overline{X}_M(\omega) - q_\alpha \frac{\sigma_X}{\sqrt{M}}, \overline{X}_M(\omega) + q_\alpha \frac{\sigma_X}{\sqrt{M}}\right],$$

then

$$\mathbb{P}\{\omega : m_X \in J_{\alpha,M}(\omega)\}) = \mathbb{P}\left(\sqrt{M}\frac{|\overline{X}_M - m_X|}{\sigma_X} \leqslant q_\alpha\right) \xrightarrow{M\to\infty} \mathbb{P}(|\mathcal{N}(0;1)| \leqslant q_\alpha) = \alpha.$$

- $J_{\alpha,M}$ is asymptotically a confidence interval for $\overline{X}_M$ at level α.
- Rate of convergence in the *CLT*

Theorem (Berry-Essen (and Shiryaev))

If $X_1 \in L^3(\mathbb{P})$, $\sigma_X > 0$, the rate in the CLT is ruled by $\forall\, n \geqslant 1$, $\forall\, x \in \mathbb{R}$,

$$\left|\mathbb{P}(\sqrt{M}(\bar{X}_M - m_X) \leqslant x\sigma_X) - \Phi_0(x)\right| \leqslant \frac{C\,\mathbb{E}|X_1 - \mathbb{E}X_1|^3}{\sigma_X^3\sqrt{M}} \frac{1}{1+|x|^3}.$$

- It turns out to be again $1/\sqrt{M}$.

Confidence interval II

- In practice the standard deviation σ_X is unknown. But

$$\begin{aligned}\overline{V}_M &= \frac{1}{M-1}\sum_{k=1}^{M}(X_k-\overline{X}_M)^2\\ &= \frac{1}{M-1}\sum_{k=1}^{M}X_k^2-\frac{M}{M-1}\overline{X}_M^2\\ &\stackrel{M\to+\infty}{\longrightarrow} \mathbb{E}\,X^2-\big(\mathbb{E}\,X\big)^2=\mathrm{Var}(X)=\sigma_X^2\end{aligned}$$

by applying twice the *SLLN*.

- Why $M-1$ and not M ? To be unbiased *i.e.* $\mathbb{E}\,V_M=\sigma_X^2$! We don't care in Numerical Probability.
- Confidence interval:

$$I_{\alpha,M}:=\left[\overline{X}_M(\omega)-q_\alpha\sqrt{\frac{\bar{V}_M}{M}},\overline{X}_M(\omega)+q_\alpha\sqrt{\frac{\bar{V}_M}{M}}\right]\simeq J_{\alpha,M}.$$

Monte Carlo simulation I

Definition (Monte Carlo simulation I)

Simulate $X_1, \ldots, X_M$.

- Define a confidence level $\alpha \in (0,1)$,
- Compute $\overline{X}_M$,
- Compute $\overline{V}_M$,
- Design the (empirical) confidence interval

$$I_{\alpha,M} = \left[\overline{X}_M(\omega) - q_\alpha \sqrt{\frac{\bar{V}_M}{M}}, \overline{X}_M(\omega) + q_\alpha \sqrt{\frac{\overline{V}_M}{M}}\right]..$$

In fact one also has that

$$\mathbb{P}\big(\{\omega : m_X \in I_{\alpha,M}(\omega)\}\big) \stackrel{M\to+\infty}{\longrightarrow} \mathbb{P}(|\mathcal{N}(0;1)| \leqslant q_\alpha) = \alpha.$$

Monte Carlo simulation II

- It is just another point of view
- Starting from the obvious

$$q_\alpha \sqrt{\frac{\overline{V}_M}{M}} = \varepsilon \Longleftrightarrow M = q_\alpha^2 \frac{\overline{V}_M}{\varepsilon^2} \simeq q_\alpha^2 \frac{\sigma_X^2}{\varepsilon^2}$$

Definition (Monte Carlo simulation II)

- Define an *RMSE* $\varepsilon > 0$ (quadratic error).
- Define a confidence level α.
- Simulate $X_1, \ldots, X_M$ and compute $\overline{X}_M$ and $\overline{V}_M$ until $\frac{\overline{V}_M}{M} = \frac{\varepsilon^2}{q_\alpha^2}$.
- Compute the confidence interval.

$$I_{\alpha,M} = [\overline{X}_M - \varepsilon, \overline{X}_M + \varepsilon].$$