MÉTHODE DE MONTE CARLO.

Alexandre Popier

Le Mans Université

2022–2023

PLAN DU COURS

BUT : CALCUL D'ESPÉRANCE.

Soient X une v.a. et $f : \mathbb{R} \to \mathbb{R}$ une fonction.

BUT

Calculer numériquement $\mathbb{E}(f(X))$.

EXEMPLES

- FINANCE : prix d'une option d'achat
  - modèle de Cox-Ross-Rubinstein :

$$C = \frac{1}{(1+r)^N} \mathbb{E}\left[\left(S_0 \prod_{i=1}^{N} T_i - K\right)^+\right].$$

$$\mathbb{P}(T_i = 1+u) = p = 1 - \mathbb{P}(T_i = 1+d).$$

  - modèle de Black-Scholes :

$$C = e^{-rT}\mathbb{E}\left[\left(S_0 e^{(r-\sigma^2/2)T + \sigma W_T} - K\right)^+\right].$$

BUT : CALCUL D'ESPÉRANCE.

Soient X une v.a. et $f : \mathbb{R} \to \mathbb{R}$ une fonction.

BUT

Calculer numériquement $\mathbb{E}(f(X))$.

EXEMPLES

- FINANCE : prix d'une option d'achat
- ASSURANCE : calcul de la prime
  - Prime pure : $\mathbb{E}(X)$,
  - Prime exponentielle : $\frac{1}{c} \ln \mathbb{E}(e^{cX})$,
  - Prime quantile : $F_X^{-1}(1-\varepsilon)$,
- CALCUL DE PERTE ou Value At Risk en finance : $\mathbb{P}(X < \text{seuil})$.

BUT : CALCUL D'ESPÉRANCE.

Soient X une v.a. et $f : \mathbb{R} \to \mathbb{R}$ une fonction.

BUT

Calculer numériquement $\mathbb{E}(f(X))$.

EXEMPLES

- FINANCE : prix d'une option d'achat
- ASSURANCE : calcul de la prime
- CALCUL DE PERTE ou Value At Risk en finance : $\mathbb{P}(X < \text{seuil})$.
- CALCUL D'INTÉGRALE : $\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(u)g(u)du$
- ...

CALCUL DU COÛT DE RÉASSURANCE TEMPÊTE :

Tarification par les méthodes probabilistes

À l'aide de l'ajustement des coûts et fréquence, **deux matrices** (une pour chaque modélisation) contenant chacune **100 000 années de sinistralité fictives** ont été crées.

	V1	V2	V3	V4	V5
1	6 689 071 €	10 663 857 €	11 545 282 €	13 560 650 €	7 310 391 €
2	21 020 782 €	7 387 657 €	0 €	0 €	0 €
3	8 056 134 €	6 246 839 €	8 277 316 €	11 710 391 €	18 754 958 €
4	8 000 555 €	7 162 079 €	6 662 135 €	6 987 227 €	14 444 597 €
5	10 361 366 €	10 700 414 €	6 586 416 €	6 786 449 €	15 219 681 €

Puis, grâce à la **fonction XS** et à la **somme des coûts** restants par ligne, on peut obtenir les coûts fictifs annuel à la **charge du réassureur.**

	Charge réassureur annuelle
1	18 549 303 €
2	11 020 782 €
3	17 274 496 €
4	6 229 536 €
5	25 000 000 €

16 — MODÉLISATION DE SINISTRES EXTRÊMES PAR L'ESTIMATEUR DE HILL POUR LA TARIFICATION D'UNE TRANCHE XS 25/08/2022

PLAN

1. MÉTHODE DE MONTE CARLO
   - Loi des grands nombres
   - Contrôle de l'erreur par TCL
   - Remarques importantes
2. PROBLÈME DE SIMULATION
   - Théorème fondamental
   - Simulation de la loi uniforme
   - Méthode d'inversion
3. VÉRIFICATIONS

PLAN

1. MÉTHODE DE MONTE CARLO
   - Loi des grands nombres
   - Contrôle de l'erreur par TCL
   - Remarques importantes
2. PROBLÈME DE SIMULATION
   - Théorème fondamental
   - Simulation de la loi uniforme
   - Méthode d'inversion
3. VÉRIFICATIONS

THÉORÈME

Soit $(Y_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires.

HYPOTHÈSES

- indépendance
- distribution identique (comme une v.a. Y)
- $\mathbb{E}(|Y|) < +\infty$.

Alors presque sûrement :

$$\lim_{N \to +\infty} \overline{Y_N} = \lim_{N \to +\infty} \frac{1}{N}(Y_1 + \ldots + Y_N) = \mathbb{E}(Y).$$

Autrement dit, pour N assez grand

$$\frac{1}{N} \sum_{i=1}^{N} Y_i \approx \mathbb{E}(Y).$$

Méthode de Monte Carlo.

Méthode

Pour calculer $\mu = \mathbb{E}(f(X))$,

- *simuler* N *v.a.* $(X_i)_{1 \leq i \leq N}$ *i.i.d. de même loi que* X,
- *poser :* $Y_i = f(X_i)$ *et*

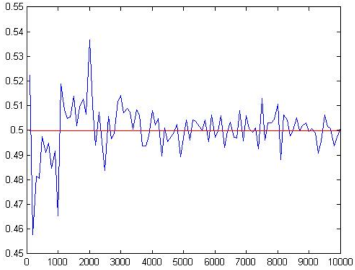
$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \frac{f(X_1) + \ldots + f(X_N)}{N}.$$

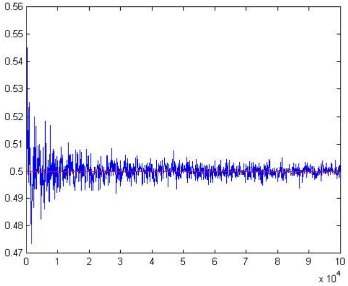
Loi des grands nombres : $\hat{\mu}_N \approx \mu$ pour N grand.

Problème : quelle est l'erreur commise ?

LOI EXPONENTIELLE (2) (ESPÉRANCE 1/2)

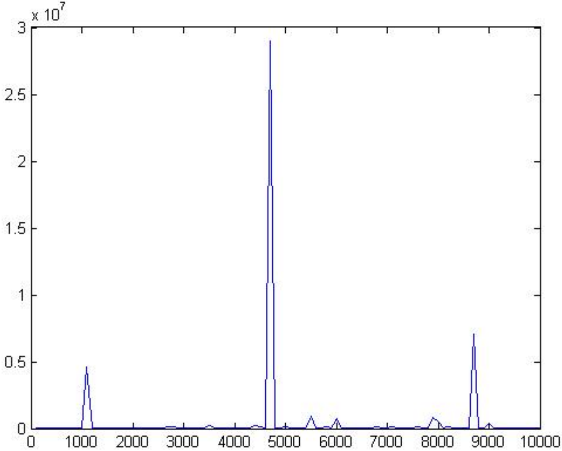
TRACÉ DE $\hat{\mu}_N$ EN FONCTION DE N :





Loi de Pareto (0,5) pas d'espérance

Tracé de $\hat{\mu}_N$ en fonction de N :



Plan

Théorème

Soit $(Y_i)_{i \in \mathbb{N}}$ une suite de v.a.

Hypothèses

- indépendance
- distribution identique (comme une v.a. Y)
- $\mathbb{E}(|Y|^2) < +\infty$.

Alors $\overline{Y_N} \cong \mathcal{N}(\mu, \sigma^2/N)$: pour tout $a < b$

$$\lim_{N \to +\infty} \mathbb{P}\left(a < \sqrt{N}\frac{\overline{Y_N} - \mu}{\sigma} < b\right) = \mathbb{P}(a < Z < b), \; Z \sim \mathcal{N}(0,1).$$

Intervalle de confiance.

TCL : $\mu = \mathbb{E}(Y_1)$, $\overline{Y}_N = \frac{1}{N}(Y_1 + \ldots + Y_N)$

$$\begin{aligned}\mathbb{P}\left(\overline{Y}_N - a\frac{\sigma}{\sqrt{N}} \leq \mu \leq \overline{Y}_N + a\frac{\sigma}{\sqrt{N}}\right) &= \mathbb{P}\left(-a \leq \sqrt{N}\frac{\overline{Y}_N - \mu}{\sigma} \leq a\right) \\ &\approx \mathbb{P}(|Z| \leq a),\end{aligned}$$

où $\sigma^2 = \mathrm{Var}(Y_1)$.
Pour un niveau de confiance $\alpha \in [0, 1]$ fixé, il existe $c_\alpha > 0$ tel que

$$\mathbb{P}(|Z| \leq c_\alpha) = \alpha.$$

Donc avec $a = c_\alpha$, pour n grand

$$\mathbb{P}\left(\mu \in I_{\alpha,N}\right) \approx \mathbb{P}(|Z| \leq c_\alpha) = \alpha$$

avec

$$I_{\alpha,N} = \left[\overline{Y}_N - c_\alpha\frac{\sigma}{\sqrt{N}}, \overline{Y}_N + c_\alpha\frac{\sigma}{\sqrt{N}}\right] \quad \text{: intervalle de confiance.}$$

Méthode

- *Simuler N v.a. $(X_i)_{1 \leq i \leq N}$ i.i.d. de même loi que X.*
- *Poser :*

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \frac{f(X_1) + \ldots + f(X_N)}{N}.$$

- *Erreur donnée par un intervalle de confiance :*

$$\mathbb{P}\left(\mu \in I_{\alpha,N}\right) \approx \alpha$$

avec

$$I_{\alpha,N} = \left[\hat{\mu}_N - c_\alpha \frac{\sigma}{\sqrt{N}}, \hat{\mu}_N + c_\alpha \frac{\sigma}{\sqrt{N}}\right], \quad \sigma^2 = Var(f(X)).$$

Problème : on ne connaît pas en général la variance σ^2.

On estime σ^2 grâce à l'estimateur $S_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y}_N)^2$. On peut montrer que

$$\mathbb{E}S_N^2 = \sigma^2 = \mathrm{Var}(Y) \text{ et } \lim_{N\to+\infty} S_N^2 = \sigma^2.$$

MÉTHODE

- *Simuler N v.a. $(X_i)_{1\leq i\leq N}$ i.i.d. de même loi que X.*
- *Poser :*

$$\hat{\sigma}_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(f(X_i) - \hat{\mu}_N)^2}.$$

MÉTHODE

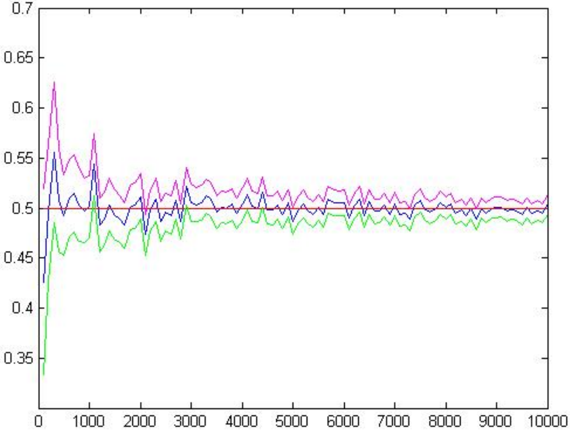
1. *Simuler N v.a. $(X_n)_{1\leq n\leq N}$ i.i.d. de même loi que X.*
2. *Poser :*

$$\begin{aligned}\hat{\mu}_N &= \frac{1}{N}\sum_{i=1}^{N} f(X_i) = \frac{f(X_1)+\ldots+f(X_N)}{N}, \quad \texttt{(mean)}\\ \hat{\sigma}_N &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(f(X_i)-\hat{\mu}_N)^2} \quad \texttt{(std)}.\end{aligned}$$

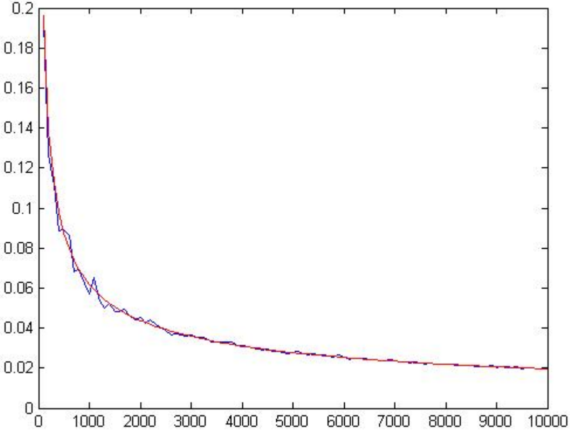
3. *Erreur donnée par intervalle de confiance avec niveau de confiance α*

$$I_{\alpha,N} = \left[\hat{\mu}_N - c_\alpha \frac{\hat{\sigma}_N}{\sqrt{N}}, \hat{\mu}_N + c_\alpha \frac{\hat{\sigma}_N}{\sqrt{N}}\right].$$

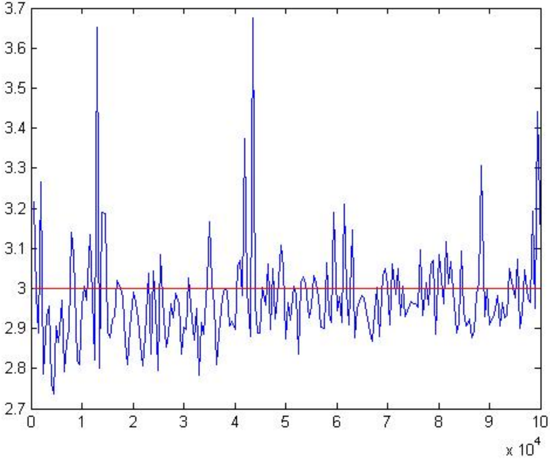
Loi exponentielle (2) (espérance 1/2)



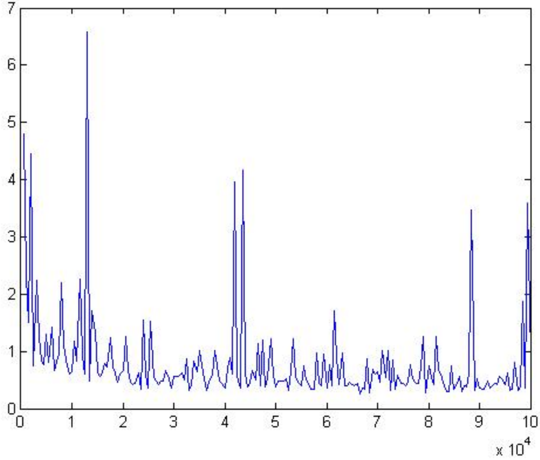
LOI EXPONENTIELLE (2) (DÉCROISSANCE EN $1/\sqrt{N}$)



PARETO (1,5) (ESPÉRANCE 3, PAS DE VARIANCE)



PARETO (1,5) (TAILLE INTERVALLE CONFIANCE)



Vitesse de convergence du TCL.

Théorème de Berry-Essen

Si Y_1 a un moment d'ordre 3 ($\mathbb{E}(|Y_1|^3) < +\infty$), alors il existe une constante $A > 0$ universelle telle que

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\sqrt{n} \frac{\overline{Y_n} - \mu}{\sigma} \leq x \right) - \mathcal{N}(x) \right| \leq \frac{A}{\sqrt{n}} \frac{\mathbb{E}|Y_1 - \mu|^3}{\sigma^3}.$$

- Vitesse en $1/\sqrt{n}$, optimale en générale (prendre $Y_1 = \pm 1$ avec proba 1/2).
- A pas connue de façon exacte.
- Amélioré par Katz-Petrov ou Edgeworth.

DIMENSION DU PROBLÈME ?

Si

- X vecteur aléatoire : $X = (X_1, \cdots, X_d)$
- $f : \mathbb{R}^d \to \mathbb{R}$

alors

- MMC encore valable pour calculer $\mathbb{E}(f(X))$
- Vitesse toujours en $1/\sqrt{N}$. Indépendant de d.

Si X a densité g :

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(y)g(y)dy.$$

- Méthodes déterministes : *vitesse dépend de d !*

MÉTHODE DE MONTE CARLO : REMARQUES.

OBLIGATOIRE

Donner $\hat{\mu}_N$ sans intervalle de confiance n'a aucune valeur !

ERREUR

Pour diminuer la taille de IC,

- diminuer le niveau de confiance α,
- augmenter N,
- diminuer σ ($\longrightarrow$ réduction de variance).

AVANT :

SIMULATION

Que signifie « Simuler N v.a. $(X_i)_{1 \leq i \leq N}$ i.i.d. de même loi que X » ?

PLAN

PLAN

Théorème fondamental de la simulation.

Théorème

Toute variable aléatoire X à valeurs dans $\mathbb{R}^d$ peut être simulée sous la forme

$$X \underset{\text{en loi}}{=} f(U_1, U_2, \dots, U_n)$$

où

- $(U_1, U_2, \dots, U_n)$ est uniformément répartie sur $[0, 1]^n$,
- la fonction $f : \mathbb{R}^n \to \mathbb{R}^d$ est borélienne et a ses points de discontinuité dans un ensemble Lebesgue-négligeable.

Il est même possible de réaliser ceci en imposant $n = 1$ ou $n = d$ ou encore $n \geq 1$ donné.

Remarque : f est « explicite ».

Plan

Considérons une suite finie $x := x_1, \dots, x_n \in \{0, 1\}$. Toutes les suites finies de ce type sont équiprobables et de probabilité 2^{-n}.
Certaines suites moins aléatoires que d'autres

$$1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1$$

$$1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1$$

$$1,0,0,1,0,1,1,1,0,1,0,1,1,0,0,0,1,0$$

- Quel sens donner et comment quantifier le caractère aléatoire d'une suite finie ou infinie donnée ?
- Comment produire des suites finies qui sont de « bonnes » approximations finies des suites infinies probables correspondant à des réalisations i.i.d. d'une loi donnée ?

D.H. Lehmer (1951) :

A random sequence is a vague notion... in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests traditional with statisticians...

SUITES ALÉATOIRES.

Simuler une loi de probabilité $\mathcal{L}$ consistera à mettre au point un algorithme utilisable pouvant générer des suites finies dont on considèrera que ce sont des réalisations indépendantes de loi $\mathcal{L}$.

- Méthodes **prédictibles** : ce sont des méthodes déterministes basées entièrement sur des algorithmes bien établis qui nécessitent d'être initialisés. On parlera de *suites pseudo-aléatoires*.
- Méthodes **non-prédictibles** : surtout utiles en cryptographie, où il est capital que le hasard utilisé ne soit pas prédictible ni reproductible. Elles peuvent être obtenues à partir des premières en utilisant des « fonctions de hachage », difficilement inversibles en terme de temps de calcul.

SIMULER LA LOI UNIFORME :

- produire par un algorithme des suites finies de nombres
- et les considérer comme autant de réalisations indépendantes de variables aléatoires uniformes sur $[0, 1]$.

MATHÉMATIQUEMENT :

- n sorties successives $u_1, u_2, \dots, u_n$
- Égales à $U_1(\omega), \dots, U_n(\omega)$ pour un $\omega \in \Omega$ avec $U_i : (\Omega, \mathcal{F}, \mathbb{P}) \to [0, 1]$ de loi uniforme.

SOUS QUELQUES LOGICIELS :

- Python : module `scipy.stats` et fonction `uniform.rvs`.
- R : fonction `runif`.
- Excel : fonction `ALEA`.
- VBA : fonction `Rnd`.

50 TIRAGES DE LA LOI UNIFORME

```
   ....
   ...: import numpy as np
   ...: import matplotlib.pyplot as plt
   ...: import scipy.stats as sps

[0.70288476 0.76628691 0.89642223 0.87380386 0.66681123 0.01761438
 0.43677983 0.45570226 0.41093659 0.5460631  0.85845819 0.03566242
 0.85637309 0.88448242 0.92524824 0.31051126 0.88835373 0.24872787
 0.07766097 0.03484874 0.34209719 0.99342923 0.30198955 0.70352559
 0.02886828 0.0472146  0.07076327 0.71171284 0.59176686 0.46394178
 0.48309298 0.54218383 0.42301015 0.87962042 0.84441035 0.16524449
 0.03193218 0.08557391 0.00218503 0.17321527 0.08321323 0.84451071
 0.11904083 0.00173334 0.57390688 0.18664527 0.83974943 0.15700128
 0.73346817 0.43126793]

In [2]: print(np.random.uniform(0,1,50))
```

PLAN

Définition

La fonction de répartition de X, notée F_X, est définie sur $\mathbb{R}$ par :

$$\forall z \in \mathbb{R}, \; F_X(z) = \mathbb{P}(X \leq z).$$

Proposition

Soit F_X la fonction de répartition d'une v.a.r. X. Alors :

1. $F_X(z) \in [0, 1]$.
2. F_X *est croissante.*
3. $\lim\limits_{z \to -\infty} F_X(z) = 0$ *et* $\lim\limits_{z \to +\infty} F_X(z) = 1$.
4. F_X *est continue à droite et a une limite à gauche en tout point.*

DÉFINITION

Une v.a. X est à densité (par rapport à la mesure de Lebesgue) s'il existe f t.q.

- *pour tout $z \in \mathbb{R}$, $f(z) \geq 0$;*
- $\displaystyle\int_{-\infty}^{+\infty} f(z)\,dz = 1$;
- *et pour tout $-\infty \leq a < b \leq +\infty$, $\mathbb{P}(a < X \leq b) = \int_a^b f(t)dt$.*

EXEMPLE : LOI UNIFORME SUR $[a, b]$

X suit une loi uniforme sur $[a, b]$ si sa densité f est

$$f(z) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(z).$$

PROPOSITION

Si X a pour densité f, alors

- $\forall z \in \mathbb{R},\ F_X(z) = \mathbb{P}(X \leq z) = \int_{-\infty}^{z} f(t)dt$.
- F_X *est continue.*
- F_X *est dérivable aux points de continuité de f avec* $f_X(z) = F'_X(z)$.

EXEMPLE : LOI UNIFORME SUR $[a, b]$

Si X suit la loi uniforme sur $[a, b]$, alors

- $F_X(z) = 0$ si $z \leq a$,
- $F_X(z) = \dfrac{z-a}{b-a}$ si $z \in [a, b]$,
- $F_X(z) = 1$ pour $z \geq b$.

DÉFINITION

Une v.a. X est discrète si elle ne prend qu'un nombre fini (ou dénombrable) de valeurs $\{x_i \in \mathbb{R},\ i \in \mathbb{N}\}$ avec probabilité $p_i = \mathbb{P}(X = x_i) \geq 0$. De plus $\sum_{i=0}^{+\infty} \mathbb{P}(X = x_i) = \sum_{i=0}^{+\infty} p_i = 1$.

REPRÉSENTATION en tableau

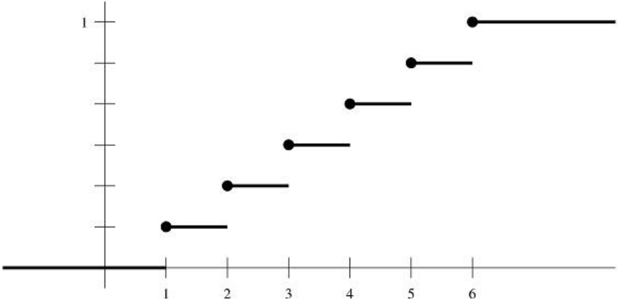
X	x_0	x_1	x_2	x_3	$\ldots$	(valeurs prises parX)
$\mathbb{P}(X = x_i)$	p_0	p_1	p_2	p_3	$\ldots$	(probabilité)

PROPOSITION

Si X est une v.a. discrète, F_X est constante par morceaux.

EXEMPLE : DÉ À SIX FACES.

FONCTION DE RÉPARTITION d'une v.a. de loi uniforme sur $\{1, \ldots, 6\}$



MÉTHODE D'INVERSION.

Si X est une v.a.r. alors la fonction de répartition F_X est définie par

$$\forall z \in \mathbb{R}, \; F_X(z) = \mathbb{P}(X \leq z).$$

Comme F_X est croissante, on peut définir la fonction pseudo-inverse q_X de F_X ainsi :

$$\forall u \in (0, 1), \; q_X(u) = \inf\{z \in \mathbb{R}, \; F_X(z) > u\}.$$

THÉORÈME

Si U suit une loi uniforme sur $[0, 1]$, $q_X(U)$ suit la même loi que X.

PROPOSITION

Si F est inversible, alors $q = F^{-1}$.

EXEMPLES.

Si U suit une loi uniforme sur $[0,1]$, alors

LOI UNIFORME SUR $[a,b]$

$X = a + (b-a)U$ suit la loi uniforme sur $[a,b]$.

LOI EXPONENTIELLE

$X = -\frac{1}{\lambda}\ln(1-U)$ suit la loi exponentielle de paramètre λ.

LOI DE CAUCHY

$X = c\tan(\pi(U-1/2))$ suit la loi de Cauchy de paramètre c.

LOI DE BERNOULLI

Si $U < 1-p$, $X = 0$, sinon $X = 1$: suit la loi de Bernoulli de paramètre $p \in [0,1]$.

PLAN

1. MÉTHODE DE MONTE CARLO
   - Loi des grands nombres
   - Contrôle de l'erreur par TCL
   - Remarques importantes
2. PROBLÈME DE SIMULATION
   - Théorème fondamental
   - Simulation de la loi uniforme
   - Méthode d'inversion
3. VÉRIFICATIONS

QUELLES VÉRIFICATIONS ?

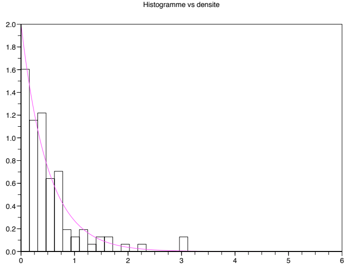
DEUX TYPES de vérifications :

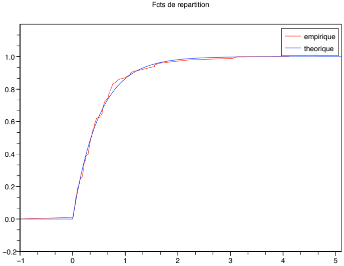
- histogramme normalisé contre densité (version fonctionnelle de la LGN)
- fonctions de répartition empirique contre théorique (Kolmogorov-Smirnov).

Dans les deux cas, il faut d'abord simuler un grand nombre de fois la loi en question. Pour $n \in \mathbb{N}^*$, on obtient un vecteur $(x_1, \dots, x_n)$ qui contient n réalisations de la loi de X.

Exemple sur la loi exponentielle(2).

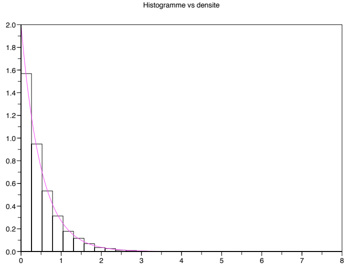
N=100

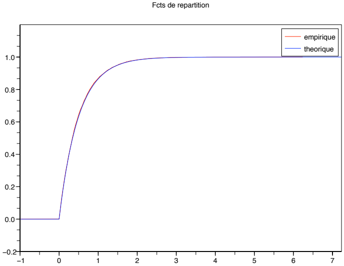




Exemple sur la loi exponentielle(2).

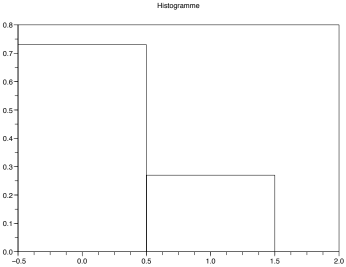
N=10000

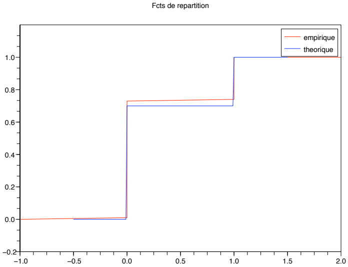




Exemple sur la loi de Bernoulli(0.3).

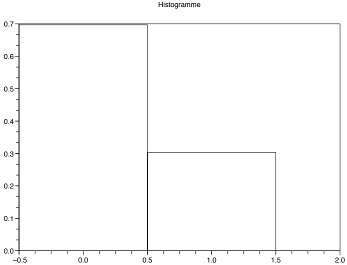
n=100

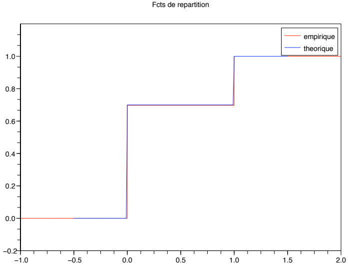




Exemple sur la loi de Bernoulli(0.3).

n=10000





Erreur de simulation !