

Méthodes numériques en probabilités et statistique
A. Gloter et A. Guilloux

Organisation

- ▶ Les documents (slides des cours, codes python, etc) sont sur ma pageweb <http://www.math-evry.cnrs.fr/members/aguilloux/welcome>
- ▶ Pour me joindre agathe.guilloux@univ-evry.fr
- ▶ Le cours est en 3 parties
 - ▶ Partie 1 (A. Guilloux, 2 séances) : Simulations de v.a. et méthode de Monte Carlo
 - ▶ Partie 2 (A. Gloter, 3 séances) : Chaînes de Markov et MCMC
 - ▶ Partie 3 (A. Guilloux, 1 séance) : Bootstrap
- ▶ Le contrôle continu se fera en 2 parties : un DM (avec code) sur la partie 1, un contrôle d'une heure pour la partie 2
- ▶ Vous devez installer python 3 avec anaconda distribution (> 4) voir <https://www.anaconda.com/products/individual>

Chapitre 1 : Introduction

Introduction générale

Fonctions de répartition, fonctions quantiles : vraies et empiriques

Générateurs congruentiels

Chapitre 2: Simulation de variables aléatoires

Méthode d'inversion

Méthode de rejet

Algorithme de Box et Muller pour les gaussiennes

Méthode Monte-Carlo

Principe

Importance sampling

Chapitre 1 : Introduction

Introduction générale : notations

- ▶ On note X une variable aléatoire (v.a.) , i.e. une application mesurable de $(\Omega, \mathcal{F}, \mathbb{P})$ dans (E, \mathcal{E}) .
- ▶ On va considérer presque partout que $E = \mathbb{R}$ et $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$
- ▶ Dans ce cas, on notera F la fonction de répartition (f.d.r.) de cette v.a. et, si elle existent, f sa densité, $\mathbb{E}(X)$ et $\mathbb{V}(X)$ ses espérance et variance.

Introduction générale : échantillon et réalisation

- ▶ En probabilités, comme en statistique, on considère souvent des échantillons X_1, \dots, X_n constitués de v.a. indépendantes et identiquement distribuées (i.i.d.) on dit aussi que ce sont des copies i.i.d. de X
- ▶ On va chercher dans cette partie à obtenir pour une f.d.r. F donnée une réalisation

$$x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$$

de X_1, \dots, X_n

- ▶ cela servira à illustrer certains résultats de probabilités, faire des calculs numériques (intégrales, etc), cela sert aussi en cryptographie par exemple.

Vraie fonction de répartition

Définition

Une fonction F de \mathbb{R} dans $[0, 1]$ est une fonction de répartition si elle est croissante, a pour limite 0 en $-\infty$, 1 en ∞ et est continue à droite avec limites à gauche.

X a pour f.d.r. F si et seulement si pour tout $x \in \mathbb{R}$: $\mathbb{P}(X \leq x) = F(x)$.

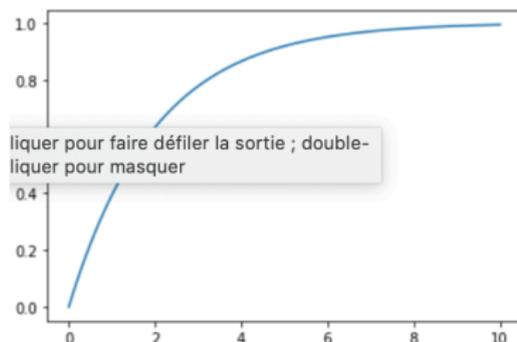


Figure 1: f.d.r de la loi $\mathcal{E}(2)$

Vraie fonction quantile

Définition

Soit F une fonction de répartition, on définit la fonction quantile associée F^- de $]0, 1[$ dans \mathbb{R} comme l'inverse généralisée de F , i.e. pour tout $u \in]0, 1[$

$$F^-(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}.$$

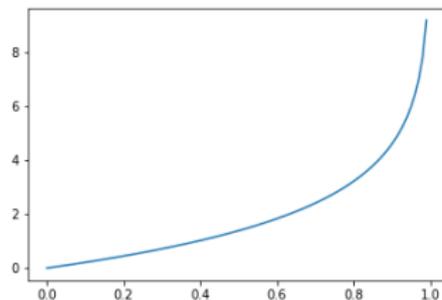


Figure 2: f.d.r de la loi $\mathcal{E}(2)$

Remarque : Si F^{-1} existe alors $F^- = F^{-1}$.

Lien entre la f.d.r. et la fonction quantile

Propriété

Soit F une f.d.r. et F^- sa fonction quantile, on a

- ▶ F^- est croissante et continue à gauche,
- ▶ et pour tous $x \in \mathbb{R}$ et $p \in]0, 1[$:

$$F(x) \geq p \iff x \geq F^-(p) \quad (1)$$

- ▶ pour tout $p \in]0, 1[$

$$F(F^-(p)) \geq p$$

avec égalité si F est continue en $F_-(p)$

- ▶ pour tout $x \in \mathbb{R}$

$$F^-(F(x)) \leq x$$

avec égalité si F^- est continue en $F(x)$

Preuve de (1)

Fonction de répartition empirique (1)

On a maintenant un échantillon i.i.d. X_1, \dots, X_n de v.a. réelles

Définition

La fonction de répartition empirique F_n associée à l'échantillon X_1, \dots, X_n est définie pour tout $x \in \mathbb{R}$ par

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i \leq x)}.$$

Pour une réalisation $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ la fonction

$F_n^\omega(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i(\omega) \leq x)}$ est une fonction de répartition constante par morceaux.

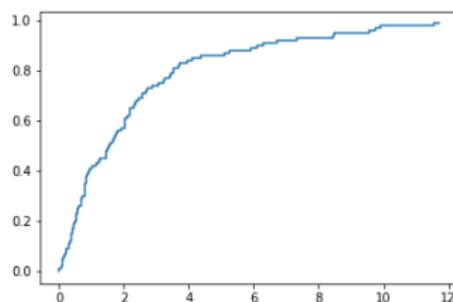


Figure 3: f.d.r empirique

Fonction de répartition empirique (2)

Théorème de Glivenko-Cantelli

Si X_1, \dots, X_n sont i.i.d. de f.d.r F alors

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s.} 0.$$

Preuve : cf. cours de Proba

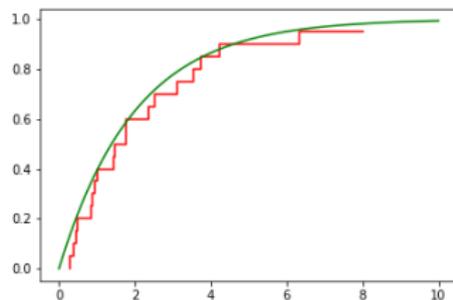


Figure 4: f.d.r vraie et empirique $n = 20$

Fonction de répartition empirique (3)

Théorème de Glivenko-Cantelli

Si X_1, \dots, X_n sont i.i.d. de f.d.r F alors

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s.} 0.$$

Preuve : cf. cours de Proba

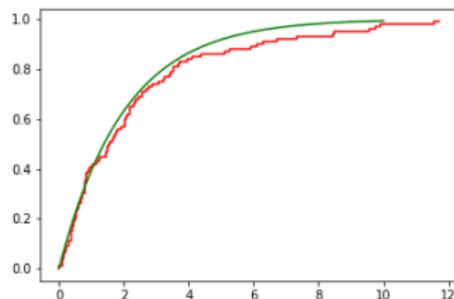


Figure 5: f.d.r vraie et empirique $n = 100$

Fonction quantile empirique (1)

Définition

La fonction quantile empirique F_n^- associée à l'échantillon X_1, \dots, X_n est définie comme l'inverse généralisée de F_n , i.e. pour tout $p \in]0, 1[$ par

$$F_n^-(u) = \inf \left\{ x \in \mathbb{R}, F_n(x) \geq u \right\}.$$

Propriété

Pour tout $p \in]0, 1[$, on a

$$F_n^-(u) = X_{(\lceil nu \rceil)}$$

où $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ sont les statistiques d'ordre de l'échantillon X_1, \dots, X_n .

Preuve

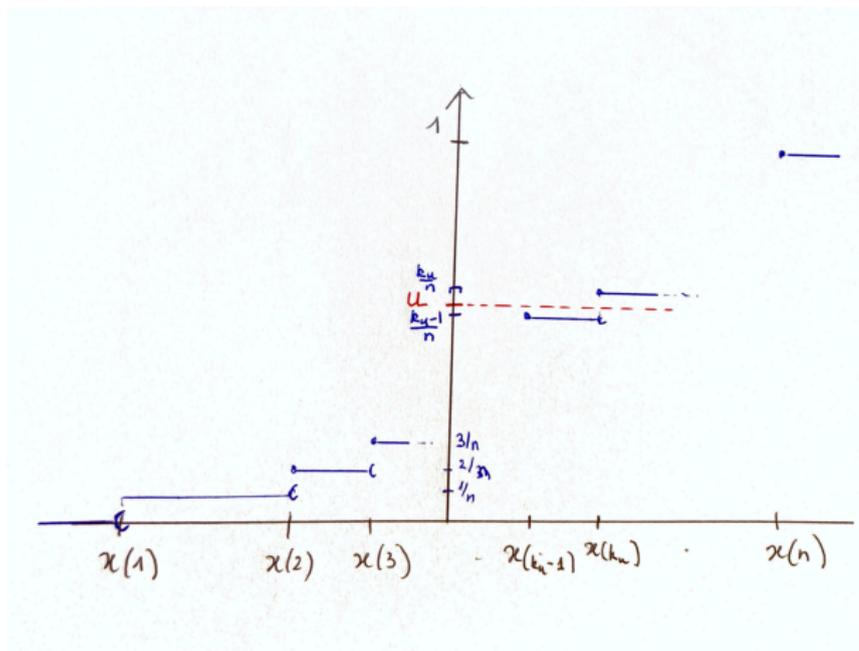


Figure 6: F_n^ω et quantiles empiriques

Preuve

Fonction quantile empirique (2)

Propriété

Pour tout $p \in]0, 1[$, si F^- est continue en p , on a

$$F_n^-(p) \xrightarrow{p.s.} F^-(p).$$

Preuve via le théorème de Glivenko-Cantelli et la continuité de F^- en p .

Si X_1, \dots, X_n sont i.i.d. de f.d.r. F on doit donc avoir pour tout $k \in \llbracket 1, n-1 \rrbracket$

$$F_n^-(k/n) = X_{(k)} \simeq F^-(k/n).$$

Quantile-quantile plot - QQ-plot

Pour vérifier si X_1, \dots, X_n ont la f.d.r. H , on représente le nuage de points

$$\left(X_{(k)}, H^{-}(k/n) \right)_{k \in [1, n-1]}$$

si H est la f.d.r., les points doivent s'aligner.

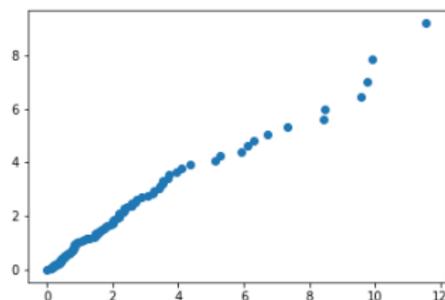


Figure 7: Réalisations d'une $\mathcal{E}(2)$ et quantiles de la $\mathcal{E}(2)$

Quantile-quantile plot - QQ-plot

Pour vérifier si X_1, \dots, X_n ont la f.d.r. H , on représente le nuage de points

$$\left(X_{(k)}, H^-(k/n) \right)_{k \in [1, n-1]}$$

si H est la f.d.r., les points doivent s'aligner.

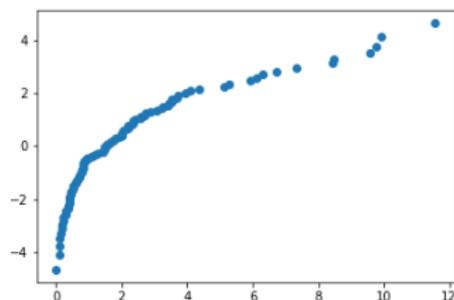


Figure 8: Réalisations d'une $\mathcal{E}(2)$ et quantiles de la $\mathcal{N}(0, 1)$

Générateurs congruentiels

Les générateurs de nombres aléatoires programmés dans la plupart des bibliothèques (c'est le cas dans `numpy`) sont de type congruentiel : ils renvoient les termes d'une suite $(z_n/m)_{n \geq 0}$ définie par

$$\begin{cases} z_0 & \text{seed ou racine} \\ z_{n+1} = az_n + c & \text{(modulo } m) \end{cases}$$

Cette suite imite très bien le hasard uniforme si la période est très (très très) grande.

Minimal standard

On prend $m = 2^{31} - 1$ et $a = 7^5$, la période est de $2^{31} - 1 \simeq 2 \times 10^9$

Mersenne twister - Matsumoto et Nishimura (1997)

L'algorithme a été amélioré pour atteindre la période de $2^{19937} - 1$.

Chapitre 2: Simulation de variables aléatoires

Introduction

On se donne une f.d.r. F (ou d'une loi) et on souhaite obtenir des réalisations x_1, \dots, x_n de X_1, \dots, X_n de f.d.r. F - on dira simuler selon F .

Suivant la forme de F , l'un ou l'autre des méthodes que nous allons voir sera la plus judicieuse.

Propriété fondamentale

Théorème

Soient U une v.a. de loi $\mathcal{U}_{]0,1[}$ et une f.d.r. F alors $F^{-}(U)$ a pour f.d.r. F

La méthode d'inversion marche si on dispose d'une forme explicite pour F^{-} : loi exponentielle, de Weibull, lois discrètes, etc

MAIS ne marche pas autrement : loi normale, loi de Student, loi du X^2 , loi gamma...

Preuve

Exercice

Loi exponentielle

Appliquer cette méthode à la loi exponentielle $\mathcal{E}(\lambda)$, pour $\lambda > 0$.

Principe

On se donne une f.d.r. F pour laquelle on souhaite obtenir des réalisations x_1, \dots, x_n de X_1, \dots, X_n de f.d.r. F et on suppose maintenant qu'elle admet une densité f

$$F(x) = \int_{-\infty}^x f(u) du \text{ pour tout } x \in \mathbb{R}.$$

On choisit une densité g (suivant laquelle on sait déjà simuler) telle qu'il existe une constante $c > 0$ et

$$f(x) \leq cg(x) \text{ pour tout } x \in \mathbb{R}.$$

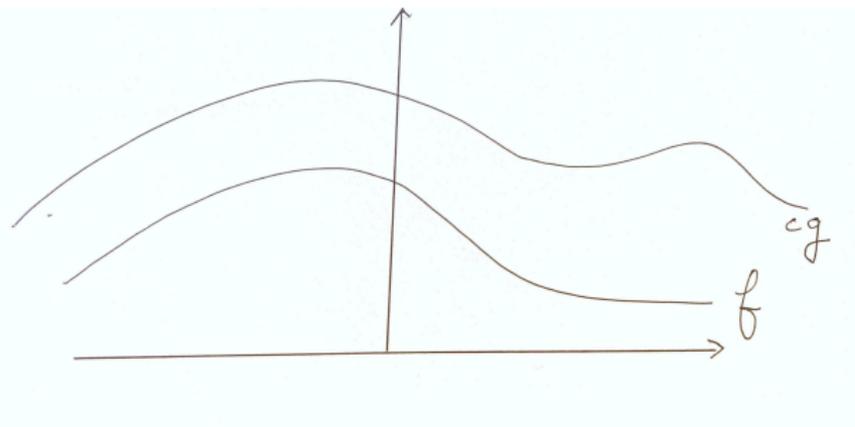


Figure 9: f et cg

Algorithme

Théorème

Soient :

- ▶ $(Y_n)_{n \geq 0}$ suite de v.a. i.i.d. de densité g indépendante de
- ▶ $(U_n)_{n \geq 0}$ suite de v.a. i.i.d. de loi $\mathcal{U}_{]0,1[}$.

On note $\tau = \inf \left\{ n \geq 1, U_n c g(Y_n) \leq f(Y_n) \right\}$ alors Y_τ a pour densité f .

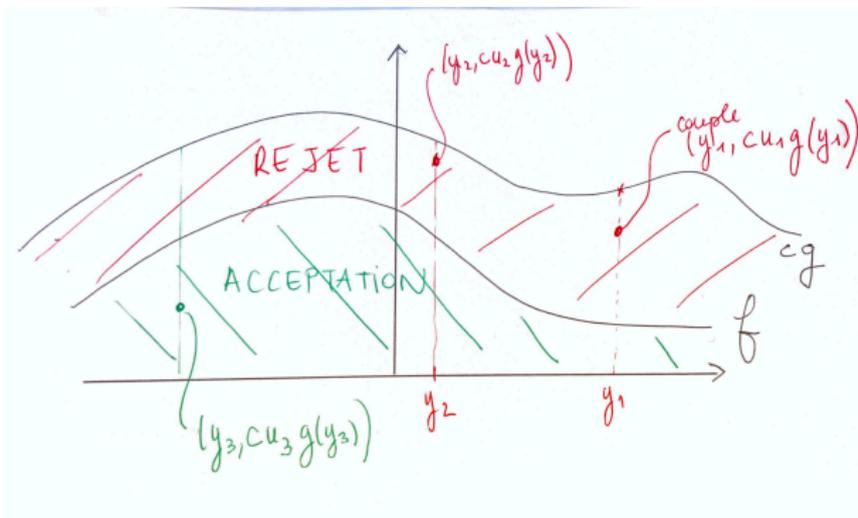


Figure 10: f et cg

Preuve

Preuve

Preuve

Simulation suivant la loi gaussienne

Algorithme de Box et Muller - 1958

Soient deux v.a. $U \sim \mathcal{U}_{]0,1[}$ et $V \sim \mathcal{E}(1)$ indépendantes, on définit

$$\begin{cases} X = \sqrt{2V} \cos(2\pi U) \\ Y = \sqrt{2V} \sin(2\pi U). \end{cases}$$

X et Y sont deux v.a. indépendantes de loi $\mathcal{N}(0, 1)$

Remarque

On peut obtenir un couple gaussienne de loi $\mathcal{N}(\mu, \Sigma)$ en translatant (X, Y) de μ et en utilisant la décomposition de Cholesky : Σ est définie positive donc il existe une matrice triangulaire inférieur L telle que $LL^T = \Sigma$.

Le vecteur $L(X, Y)^T + \mu$ a la loi $\mathcal{N}(\mu, \Sigma)$.

Preuve

Preuve

Méthode Monte-Carlo

Introduction et rappels de probabilités

C'est une méthode utilisant la loi des grands nombres qui sert à approximer une valeur numérique. Son nom est dû aux jeux des casinos de Monte-Carlo, elle a été introduite par N. Metropolis (1949).

Loi des grands nombres

Soit $(Z_n)_{n \geq 1}$ une suite de v.a. i.i.d. intégrables ($\mathbb{E}(|Z_1|) < \infty$) on a

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i \xrightarrow{p.s.} \mathbb{E}(Z_1).$$

Inégalité de Bienaymé-Tchebychev

Soit $(Z_n)_{n \geq 1}$ une suite de v.a. i.i.d. admettant une variance ($\mathbb{E}(|Z_1|^2) < \infty$) on a

$$\mathbb{P}\left(|\bar{Z}_n - \mathbb{E}(Z_1)| > \epsilon\right) \leq \frac{\mathbb{V}(Z_1)}{n\epsilon}.$$

Principe de la méthode (1)

On souhaite approximer l'intégrale

$$I = \int_{\Delta} g,$$

où g est une fonction de \mathbb{R}^d dans \mathbb{R} que l'on suppose intégrable sur Δ ($\int_{\Delta} |g| < \infty$).

On remarque pour cela que l'on peut toujours récrire l'intégrale I de la façon suivante :

$$I = \int_{\Delta} \frac{g}{f_X} f_X = \int \frac{g}{f_X} f_X$$

où f_X est une densité sur \mathbb{R}^d de support Δ (pour $x \in \mathbb{R}^d$, $f_X(x) > 0 \Leftrightarrow x \in \Delta$).

Principe de la méthode (1)

Méthode de Monte-Carlo

Si X est une v.a. de \mathbb{R}^d de densité f_X , on peut alors écrire :

$$I = \mathbb{E}\left(\frac{g}{f_X}(X)\right).$$

Soient $(X_n)_{n \geq 1}$ une suite v.a. i.i.d. de même loi que X , on approxime I par

$$\bar{I}_n = \frac{1}{n} \sum_{i=1}^n \frac{g}{f_X}(X_i).$$

\bar{I}_n converge p.s. vers I

Preuve

Illustration pour approximer π (1)

On pose

$$I = 4 \int_{[-1,1]^2} \mathbb{1}_{(x^2+y^2 \leq 1)} dx dy = \pi.$$

On considère U et V deux v.a. indépendantes de loi uniforme sur $[-1, 1]$, on peut récrire I comme

$$I = \mathbb{E}(\mathbb{1}_{(U^2+V^2 \leq 1)}).$$

On peut alors l'approximer à partir de copies i.i.d. de U et V

$$\bar{I}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(U_i^2+V_i^2 \leq 1)}$$

Illustration pour π (2)

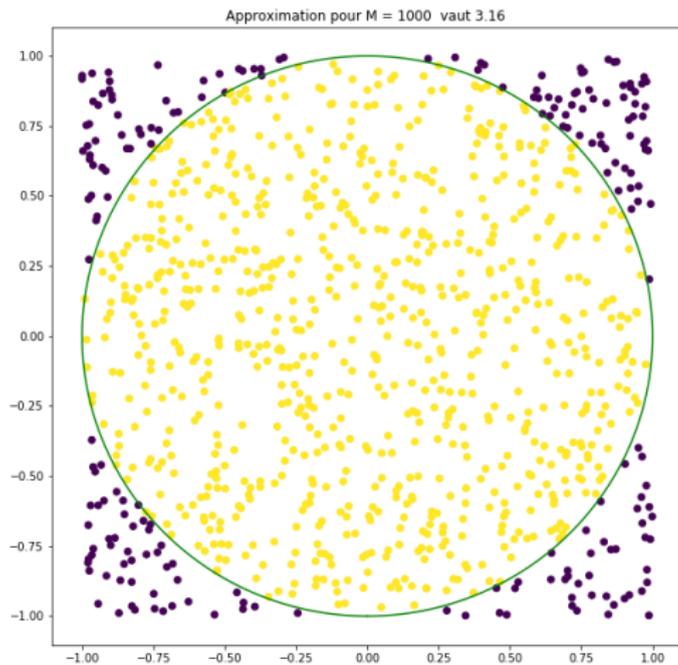
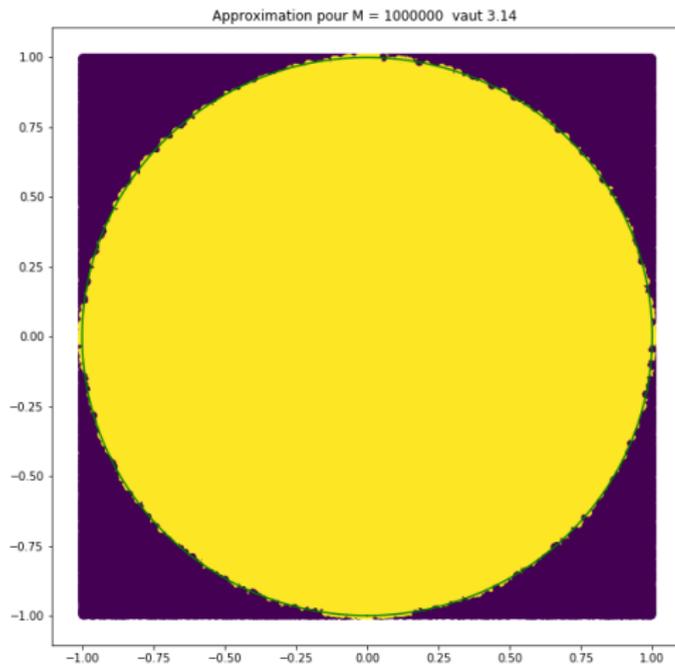


Illustration pour π (3)



Comment choisir f_X ?

On suppose maintenant que l'inégalité de Bienaymé-Tchebychev est vérifiée (i.e. $\int g^2/f_X < \infty$), on a alors

$$\mathbb{P}\left(|\bar{I}_n - I| > \epsilon\right) \leq \frac{\mathbb{V}(g/f_X(X))}{n\epsilon}.$$

Donc, à ϵ fixé, plus $\mathbb{V}(g/f_X(X))$ est petite, plus petite sera l'erreur d'approximation !

On calcule facilement

$$\mathbb{V}(g/f_X(X)) = \int_{\Delta} \frac{g^2(x)}{f_X(x)} dx - I^2$$

c'est le premier terme qu'on va chercher à rendre petit.

Exemple avec un extrême gaussien (1)

Soit $X \sim \mathcal{N}(0, 1)$, on note φ sa densité. On veut approximer

$$I = \mathbb{P}(X > 5).$$

On récrit

$$I = \mathbb{P}(X > 5) = \mathbb{E}(\mathbb{1}_{(X>5)}) = \int \mathbb{1}_{(x>5)} \varphi(x) dx = \int \frac{\mathbb{1}_{(x>5)} \varphi(x)}{\varphi(x)} \varphi(x) dx.$$

Avec les notations précédentes, on a

$$g(x) = \mathbb{1}_{(x>5)} \varphi(x) \text{ et } f_X(x) = \varphi(x) \text{ soit } \frac{g}{f_X}(x) = \mathbb{1}_{(x>5)}$$

On peut donc approximer I par

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i > 5)} \text{ avec } (X_n)_{n \geq 1} \text{ i.i.d. } \mathcal{N}(0, 1).$$

Est ce une bonne idée ? Dans ce cas, on a

$$\int \frac{g^2(x)}{f_X(x)} dx = \int \mathbb{1}_{x>5} \varphi(x) dx$$

Exemple avec un extrême gaussien (2)

Plus généralement, on peut choisir d'approximer I à l'aide de v.a. i.i.d. de loi $\mathcal{N}(\mu, 1)$. Dans ce cas, comment choisir μ ? On note φ_μ la densité associée. On peut écrire

$$I = \int \mathbb{1}_{(x>5)} \varphi(x) dx = \int \frac{\mathbb{1}_{(x>5)} \varphi(x)}{\varphi_\mu(x)} \varphi_\mu(x) dx.$$

On peut donc approximer I par

$$\tilde{I}_\mu = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{1}_{(X_i>5)} \varphi(X_i)}{\varphi_\mu(X_i)}.$$

La variance sera alors donnée par

$$\mathbb{V}(g/\varphi(X_i)) = \int \frac{\mathbb{1}_{(x>5)} \varphi^2(x)}{\varphi_\mu(x)} - \hat{I}^2.$$

Exemple avec un extrême gaussien (3)

Pour comparer cette approximation à la précédente (notée \hat{I}), il faut comparer les deux variances d'estimation ou c'est équivalent

$$\int \mathbb{1}_{x>5} \varphi(x) dx \text{ et } \int \frac{\mathbb{1}_{(x>5)} \varphi^2(x)}{\varphi_\mu(x)} dx$$

ou bien $\varphi(x)/\varphi_\mu(x)$ pour $x > 5$:

$$\frac{\varphi(x)}{\varphi_\mu(x)} = \frac{(1/\sqrt{2\pi}) \exp(-x^2/2)}{(1/\sqrt{2\pi}) \exp(-(x-\mu)^2/2)} = \exp(\mu(\mu/2 - x))$$

Si on prend $\mu \geq 0$, il faut donc que $\mu < 10$ pour ce quotient soit plus petit que 1 pour tout $x > 5$.

Conclusion : pour tout $\mu \in]0, 10]$, l'approximation par \tilde{I}_μ est meilleure que l'approximation \hat{I} (pour la borne de Bienaymé-Tchebychev).