Algorithmes stochastiques et chaînes de Markov discrètes

Séance 2 – Méthode delta, dimension, fiabilité et simulation de variables aléatoires

M2 Finance – Université Gustave Eiffel

But de la séance : terminer le chapitre 1 en reliant précision statistique, dimension, diagnostics de fiabilité, et simulation rigoureuse des variables aléatoires utilisées dans les estimateurs.

Panorama du cours

Chapitre 1	Moyennes de Monte-Carlo, convergence, intervalles de confiance, choix de n, méthode delta, dimension, exemples appliqués, simulation par inversion et acceptation–rejet.
Chapitre 2	Réduction de variance : antithétiques, variables de contrôle, conditionnement, stratification, importance sampling, événements rares.
Chapitre 3	Espérances conditionnelles, projection L^2, régression, programmation dynamique, Longstaff–Schwartz.
Chapitre 4	Quasi-Monte-Carlo, suites à faible discrépance, Robbins–Monro, Kiefer–Wolfowitz.

Séance 2. On termine le chapitre 1 : méthode delta, dimension et coût de calcul, diagnostics de fiabilité, exemples appliqués, inversion, acceptation–rejet et simulation uniforme dans le disque.

Fil directeur de la séance

Un calcul Monte-Carlo complet suit la chaîne

$$\boxed{I} \longrightarrow \boxed{I = \mathbb{E}(Y)} \longrightarrow \boxed{\text{simulation de } Y} \longrightarrow \boxed{\widehat{I}_n} \longrightarrow \boxed{\text{erreur et diagnostics}}.$$

Les quatre questions d'aujourd'hui

1. Que devient l'erreur lorsque l'on transforme un estimateur ?
2. Pourquoi les grilles cartésiennes deviennent-elles impraticables en grande dimension ?
3. Que veut dire qu'une simulation est fiable, au-delà du simple IC affiché ?
4. Comment simuler rigoureusement les lois dont on a besoin ?

Méthode delta : énoncé

Théorème

Soit $(\widehat{\theta}_n)$ tel que

$$\sqrt{n}\,(\widehat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \tau^2),$$

et soit $\varphi : \mathbb{R} \to \mathbb{R}$ dérivable en θ. Alors

$$\sqrt{n}\,\big(\varphi(\widehat{\theta}_n) - \varphi(\theta)\big) \Longrightarrow \mathcal{N}\big(0, (\varphi'(\theta))^2\tau^2\big).$$

Idée à retenir

À l'ordre $n^{-1/2}$, on remplace φ par sa tangente :

$$\varphi(\widehat{\theta}_n) - \varphi(\theta) \approx \varphi'(\theta)(\widehat{\theta}_n - \theta).$$

Méthode delta : preuve rigoureuse

Comme φ est dérivable en θ,

$$\varphi(x) = \varphi(\theta) + \varphi'(\theta)(x-\theta) + r(x)(x-\theta), \qquad r(x) \xrightarrow[x\to\theta]{} 0.$$

En prenant $x = \widehat{\theta}_n$,

$$\sqrt{n}\left(\varphi(\widehat{\theta}_n) - \varphi(\theta)\right) = \varphi'(\theta)\sqrt{n}(\widehat{\theta}_n - \theta) + r(\widehat{\theta}_n)\sqrt{n}(\widehat{\theta}_n - \theta).$$

Or $\widehat{\theta}_n \to \theta$ en probabilité, donc $r(\widehat{\theta}_n) \to 0$ en probabilité. Comme $\sqrt{n}(\widehat{\theta}_n - \theta) = O_{\mathbb{P}}(1)$,

$$r(\widehat{\theta}_n)\sqrt{n}(\widehat{\theta}_n - \theta) = o_{\mathbb{P}}(1).$$

Conclusion

Par Slutsky, la limite est gaussienne de variance $(\varphi'(\theta))^2\tau^2$.

Exemple Bernoulli : de p vers π

Dans le calcul de π par la méthode du disque,

$$B_k = \mathbf{1}_{\{U_k \in D\}}, \qquad B_k \sim \text{Bernoulli}(p), \qquad p = \frac{\pi}{4}.$$

Alors

$$\widehat{p}_n = \frac{1}{n}\sum_{k=1}^{n} B_k, \qquad \sqrt{n}(\widehat{p}_n - p) \Longrightarrow \mathcal{N}(0, p(1-p)).$$

On veut estimer $\pi = 4p$. Avec $\varphi(p) = 4p$,

$$\sqrt{n}(\widehat{\pi}_n - \pi) \Longrightarrow \mathcal{N}\big(0, 16p(1-p)\big), \qquad \widehat{\pi}_n = 4\widehat{p}_n.$$

Un IC asymptotique à 95% est donc

$$\widehat{\pi}_n \pm 1.96 \times 4\sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}}.$$

Approximation par grille : l'idée

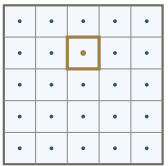
On veut approximer

$$I_d = \int_{[0,1]^d} f(x)\ \mathrm{d}x.$$

On découpe le cube en m^d petits cubes C_j de côté $1/m$. Dans chaque petit cube, on remplace $f(x)$ par sa valeur au centre x_j.

Cela donne la somme de Riemann multidimensionnelle

$$Q_m(f) = \sum_{j=1}^{m^d} f(x_j)\ \mathrm{Vol}(C_j).$$



centres des petits carrés

Comme chaque C_j a côté $1/m$,

$$\mathrm{Vol}(C_j) = \left(\frac{1}{m}\right)^d, \qquad Q_m(f) = \frac{1}{m^d} \sum_{j=1}^{m^d} f(x_j).$$

Idée à retenir

$Q_m(f)$ est l'approximation naturelle : aire du petit cube $\times$ valeur de f au centre, puis on somme sur tous les cubes.

Erreur déterministe de la grille : intuition

Supposons f L-lipschitzienne :

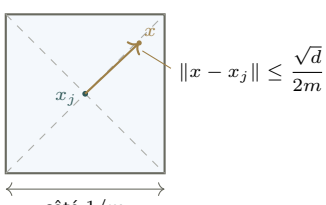
$$|f(x) - f(y)| \leq L\|x - y\|.$$

Dans un cube C_j, tous les points sont proches du centre x_j. La plus grande distance possible est la demi-diagonale :

$$\|x - x_j\| \leq \frac{\sqrt{d}}{2m}, \qquad x \in C_j.$$

Donc, dans tout le cube,

$$|f(x) - f(x_j)| \leq L\frac{\sqrt{d}}{2m}.$$



Erreur déterministe de la grille : preuve

L'erreur sur chaque petit cube vient du remplacement de $f(x)$ par $f(x_j)$:

$$I_d - Q_m(f) = \sum_{j=1}^{m^d} \int_{C_j} \big(f(x) - f(x_j)\big)\, \mathrm{d}x.$$

Donc

$$|I_d - Q_m(f)| \leq \sum_{j=1}^{m^d} \int_{C_j} |f(x) - f(x_j)|\, \mathrm{d}x.$$

Par l'hypothèse lipschitzienne et la borne sur la demi-diagonale,

$$|f(x) - f(x_j)| \leq L\frac{\sqrt{d}}{2m}, \qquad x \in C_j.$$

Ainsi

$$|I_d - Q_m(f)| \leq \sum_{j=1}^{m^d} \mathrm{Vol}(C_j)\frac{L\sqrt{d}}{2m} = \frac{L\sqrt{d}}{2m}.$$

Comme $N = m^d$, on obtient

$$|I_d - Q_m(f)| \leq \frac{L\sqrt{d}}{2} N^{-1/d}.$$

Dimension 10 : ce que coûte une grille

Supposons $d = 10$, $f : [0,1]^{10} \to [0,1]$, $L = 1$, et cherchons une précision de l'ordre de 10^{-2}.

Approximation par grille

La borne déterministe obtenue ci-dessus donne

$$|Q_m(f) - I_{10}| \leq \frac{\sqrt{10}}{2} N^{-1/10}.$$

Imposer une borne $\leq 10^{-2}$ donne

$$N \geq \left(\frac{\sqrt{10}}{2 \times 10^{-2}} \right)^{10} = 158.1^{10} \approx 10^{22}.$$

Cela correspond à environ 159 points dans chacune des dix directions.

Point de vigilance

Un nombre modéré de points par coordonnée devient astronomique lorsque l'on prend le produit cartésien.

Dimension 10 : ce que donne Monte-Carlo

Gardons le même objectif : erreur de l'ordre de 10^{-2}, avec $0 \leq f \leq 1$.

Borne probabiliste de Hoeffding

Pour Monte-Carlo,

$$\mathbb{P}\Big(\Big|\widehat{I}_N - I_{10}\Big| \geq 10^{-2}\Big) \leq 2\exp(-2N10^{-4}).$$

Pour que cette probabilité soit $\leq 5\%$, il suffit que

$$2\exp(-2N10^{-4}) \leq 0.05.$$

Donc

$$N \geq \frac{\log(40)}{2 \times 10^{-4}} \approx 1.84 \times 10^4.$$

Idée à retenir

La comparaison explique pourquoi Monte-Carlo devient intéressant en grande dimension.

Attention : les garanties ne sont pas de même nature

Les chiffres précédents ne comparent pas exactement le même type de garantie.

Grille déterministe

La borne $|Q_m(f) - I_d| \leq CN^{-1/d}$ est déterministe : si les hypothèses de régularité sont vraies, l'erreur est bornée pour cette approximation par grille.

Monte-Carlo

La borne issue de Hoeffding est probabiliste :

$$\mathbb{P}\Big(\Big|\widehat{I}_N - I_d\Big| \leq 10^{-2}\Big) \geq 95\%.$$

Elle porte sur la sortie aléatoire de l'algorithme.

Idée à retenir

Le message n'est pas que Monte-Carlo gagne toujours ; le message est que les grilles cartésiennes deviennent vite impraticables quand d augmente.

Fiabilité : que veut-on contrôler ?

À la fin d'une simulation, on observe $Y_1, \dots, Y_n$, donc $\overline{Y}_n$ et S_n. On ne voit pas directement l'erreur réelle

$$\left|\overline{Y}_n - I\right|.$$

Rôle de l'intervalle de confiance

L'IC normal

$$\overline{Y}_n \pm 1.96 \frac{S_n}{\sqrt{n}}$$

est un indicateur sérieux lorsque ses ingrédients sont stabilisés : variance finie, S_n représentatif de σ, et approximation normale déjà pertinente à la taille n utilisée.

Idée à retenir

La fiabilité consiste à vérifier que l'erreur standard affichée décrit vraiment les fluctuations actuelles de l'estimateur.

Fausse convergence : pourquoi l'IC peut être trompeur

Une **fausse convergence** se produit lorsque la moyenne semble stabilisée simplement parce que les simulations n'ont pas encore rencontré certaines valeurs rares mais importantes.

Situations à risque

- événement rare : beaucoup de zéros, puis quelques valeurs non nulles décisives ;
- queue lourde : la variance est énorme ou infinie ;
- forte asymétrie : le TCL peut être très lent ;
- variance empirique instable : S_n change encore fortement quand n augmente.

Point de vigilance

L'IC reste la bonne quantité à calculer, mais il faut vérifier que les conditions pratiques de son interprétation sont raisonnables pour la taille de simulation considérée.

Diagnostic en doublant la taille : définition

On compare la moyenne après n simulations et après $2n$ simulations :

$$\overline{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k, \qquad \overline{Y}_{2n} = \frac{1}{2n}\sum_{k=1}^{2n} Y_k.$$

Si $Y \in L^2$ de variance σ^2, alors

$$\mathrm{Var}(\overline{Y}_{2n} - \overline{Y}_n) = \frac{\sigma^2}{2n}.$$

On normalise donc par l'échelle attendue de fluctuation :

$$D_n = \frac{\overline{Y}_{2n} - \overline{Y}_n}{S_{2n}/\sqrt{2n}}.$$

Idée à retenir

D_n mesure le changement observé quand on double n, en nombre d'erreurs standards.

Diagnostic en doublant la taille : interprétation

Si l'approximation normale est raisonnable et si S_{2n} est déjà stabilisé, alors D_n doit être de l'ordre d'une normale centrée réduite.

Ce qu'on voudrait observer

- On ne demande pas $D_n \approx 0$ exactement.
- Des valeurs comme $|D_n| \leq 2$ sont compatibles avec des fluctuations normales.
- Une valeur isolée $|D_n| > 3$ est un signal d'alerte.
- Des valeurs souvent grandes lorsque l'on augmente n indiquent que l'erreur standard affichée ne décrit pas encore bien l'erreur réelle.

Remarque

Ce diagnostic ne remplace pas un théorème. Il sert à repérer les cas où l'IC asymptotique est probablement trop optimiste pour la taille de simulation actuelle.

Probabilité de ruine : contexte et risque numérique

Considérons un portefeuille d'assurance ou de crédit. Sur un an :

- L désigne la perte totale du portefeuille ;
- K désigne le capital disponible pour absorber les pertes.

La ruine correspond à $L > K$, donc

$$p = \mathbb{P}(L > K) = \mathbb{E}\left[\mathbf{1}_{\{L>K\}}\right].$$

L'estimateur naïf est

$$\widehat{p}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{L_k>K\}}.$$

Point de vigilance

Si p est petit, il peut arriver que l'échantillon ne contienne aucune ruine. Dans ce cas $\widehat{p}_n = 0$ et $S_n = 0$: l'intervalle affiché reflète seulement ce qui a été observé, pas ce qui aurait pu être observé avec une simulation beaucoup plus longue.

Erreur relative réalisée : ce que l'on voudrait connaître

Pour une quantité $I \neq 0$, l'erreur relative réalisée est

$$\frac{\left|\widehat{I}_n - I\right|}{|I|}.$$

Elle compare l'erreur effective au niveau de la quantité à estimer.

Pourquoi on ne peut pas la calculer directement

Dans un vrai problème numérique, I est précisément inconnu. Donc le numérateur $\left|\widehat{I}_n - I\right|$ est inconnu, et le dénominateur $|I|$ aussi.

Idée à retenir

On remplace donc l'erreur réalisée par une échelle typique d'erreur. C'est là que la variance intervient : l'écart-type de $\widehat{I}_n$ mesure la taille typique de $\widehat{I}_n - I$.

Erreur standard relative : version théorique et version pratique

La version théorique est

$$\text{erreur standard relative} = \frac{\sqrt{\mathrm{Var}(\widehat{I}_n)}}{|I|}.$$

Elle n'est pas directement observable, mais elle explique l'ordre de grandeur attendu de l'erreur relative réalisée.

Version estimée en pratique

Pour une moyenne Monte-Carlo,

$$\sqrt{\mathrm{Var}(\widehat{I}_n)} \approx \frac{S_n}{\sqrt{n}}, \qquad I \approx \widehat{I}_n.$$

On affiche donc souvent

$$\widehat{\mathrm{RSE}}_n = \frac{S_n/\sqrt{n}}{\left|\widehat{I}_n\right|},$$

quand $\widehat{I}_n \neq 0$.

Cas Bernoulli : retour sur la probabilité de ruine

Pour la probabilité de ruine,

$$Y = \mathbf{1}_{\{L>K\}} \sim \text{Bernoulli}(p), \qquad \widehat{p}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k.$$

La formule théorique utilise le vrai p :

$$\text{Var}(\widehat{p}_n) = \frac{p(1-p)}{n}.$$

Donc

$$\frac{\sqrt{\text{Var}(\widehat{p}_n)}}{p} = \sqrt{\frac{1-p}{np}} \simeq \frac{1}{\sqrt{np}} \qquad (p \ll 1).$$

Lecture pratique

Comme p est inconnu, on le remplace par $\widehat{p}_n$ si l'échantillon contient assez d'événements. La formule dit surtout que le bon paramètre est np : le nombre moyen d'événements observés.

Cas Bernoulli : conséquence numérique

Lorsque p est petit,

$$\text{erreur standard relative} \simeq \frac{1}{\sqrt{np}}.$$

Pour viser environ 10%, il faut

$$\frac{1}{\sqrt{np}} \approx 0.1 \quad \Longleftrightarrow \quad np \approx 100.$$

Idée à retenir

Il faut donc observer de l'ordre de cent ruines pour obtenir une précision relative d'environ 10% avec l'estimateur naïf.

Point de vigilance

Si $p = 10^{-5}$, cela signifie $n \approx 10^7$ simulations. C'est la raison pour laquelle les événements rares demandent des méthodes de réduction de variance.

Méthode d'inversion : théorème général

Soit F une fonction de répartition sur $\mathbb{R}$. On définit son inverse généralisée par

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \qquad 0 < u < 1.$$

Théorème d'inversion

Si $U \sim \mathcal{U}([0,1])$, alors

$$X = F^{-1}(U)$$

a pour fonction de répartition F.

Idée à retenir

Une variable uniforme sur $[0,1]$ suffit à fabriquer une variable de fonction de répartition F, dès que l'on sait calculer F^{-1}.

Méthode d'inversion : preuve

Il suffit de montrer que, pour tout $x \in \mathbb{R}$,

$$\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}.$$

Fixons $u \in (0, 1)$.

Équivalence clé

$$F^{-1}(u) \leq x \quad \Longleftrightarrow \quad u \leq F(x).$$

En effet, si $F^{-1}(u) \leq x$, alors par croissance et continuité à droite de F, on obtient $u \leq F(x)$. Réciproquement, si $u \leq F(x)$, alors x appartient à l'ensemble $\{t : F(t) \geq u\}$, donc son infimum est inférieur ou égal à x. Ainsi,

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Exemple : Bernoulli par inversion

Soit $X \sim \text{Bernoulli}(p)$, à valeurs dans $\{0, 1\}$. Sa fonction de répartition est

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Son inverse généralisée est donc

$$F^{-1}(u) = \begin{cases} 0, & 0 < u \leq 1 - p, \\ 1, & 1 - p < u < 1. \end{cases}$$

Si $U \sim \mathcal{U}([0, 1])$, alors

$$F^{-1}(U) = \mathbf{1}_{\{U > 1-p\}}.$$

Comme $1 - U \overset{\mathcal{L}}{=} U$, on utilise souvent la forme équivalente

$$X = \mathbf{1}_{\{U \leq p\}}.$$

Exemples continus d'inversion

Exponentielle de paramètre λ

Pour $x \geq 0$,

$$F(x) = 1 - e^{-\lambda x}.$$

On résout $u = 1 - e^{-\lambda x}$:

$$F^{-1}(u) = -\frac{1}{\lambda} \log(1-u).$$

Comme $1 - U \overset{\mathcal{L}}{=} U$,

$$X = -\frac{1}{\lambda} \log U$$

a la loi exponentielle.

Pareto

Si $F(x) = 1 - x^{-\alpha}$ sur $[1, +\infty)$, alors

$$X = (1-U)^{-1/\alpha}$$

suit une loi de Pareto.

Acceptation–rejet : théorème général

On veut simuler une densité cible f sur $\mathbb{R}^d$. On sait simuler une densité g et l'on suppose qu'il existe $c \geq 1$ tel que

$$f(x) \leq c g(x) \qquad \text{pour tout } x.$$

On répète :

1. simuler $Y \sim g$;
2. simuler $U \sim \mathcal{U}([0,1])$, indépendant de Y ;
3. accepter Y si

$$U \leq \frac{f(Y)}{c g(Y)}.$$

Résultat

La variable acceptée a pour densité f. De plus, la probabilité d'acceptation vaut $1/c$.

Acceptation–rejet : quelle loi veut-on identifier ?

Notons

$$\mathrm{Acc} = \left\{ U \leq \frac{f(Y)}{cg(Y)} \right\}.$$

L'algorithme renvoie Y, mais seulement dans les essais où Acc a lieu. La loi de la variable renvoyée est donc la loi conditionnelle

$$\mathcal{L}(Y \mid \mathrm{Acc}).$$

Pour montrer que cette loi a densité f, il suffit de montrer que, pour tout borélien B,

$$\mathbb{P}(Y \in B \mid \mathrm{Acc}) = \int_B f(y)\, \mathrm{d}y.$$

Idée à retenir

On commence donc par calculer $\mathbb{P}(Y \in B, \mathrm{Acc})$, puis on divisera par $\mathbb{P}(\mathrm{Acc})$.

Acceptation–rejet : calcul de la loi conditionnelle

On utilise la formule des probabilités totales sous forme continue : conditionner par la valeur proposée $Y = y$, puis intégrer contre la densité de Y.

Formule utilisée

Si Y a densité g, alors pour tout événement E dépendant de Y et d'un autre aléa,

$$\mathbb{P}(Y \in B, E) = \int_B \mathbb{P}(E \mid Y = y) g(y) \, \mathrm{d}y.$$

Ici

$$E = \mathrm{Acc} = \left\{ U \leq \frac{f(Y)}{c g(Y)} \right\}.$$

Donc

$$\mathbb{P}(Y \in B, \mathrm{Acc}) = \int_B \mathbb{P}\left(U \leq \frac{f(y)}{c g(y)} \mid Y = y \right) g(y) \, \mathrm{d}y.$$

Acceptation–rejet : calcul de la loi conditionnelle

Comme U est uniforme et indépendant de Y, et comme $0 \leq f(y)/(cg(y)) \leq 1$,

$$\mathbb{P}\left(U \leq \frac{f(y)}{cg(y)} \mid Y = y\right) = \frac{f(y)}{cg(y)}.$$

Ainsi

$$\mathbb{P}(Y \in B, \mathrm{Acc}) = \int_B \frac{f(y)}{cg(y)} g(y)\ \mathrm{d}y = \frac{1}{c} \int_B f(y)\ \mathrm{d}y.$$

En prenant $B = \mathbb{R}^d$, on obtient

$$\mathbb{P}(\mathrm{Acc}) = \frac{1}{c}.$$

Donc

$$\mathbb{P}(Y \in B \mid \mathrm{Acc}) = \frac{\mathbb{P}(Y \in B, \mathrm{Acc})}{\mathbb{P}(\mathrm{Acc})} = \int_B f(y)\ \mathrm{d}y.$$

Acceptation–rejet : nombre d'essais

À chaque essai,

$$p_{\text{acc}} = \mathbb{P}(\text{Acc}) = \frac{1}{c}.$$

Les essais successifs sont indépendants, car on resimule un nouveau couple (Y, U) à chaque tentative. Si N est le nombre d'essais jusqu'à la première acceptation, alors l'événement $\{N = k\}$ signifie :

$$\underbrace{\text{rejet}, \ldots, \text{rejet}}_{k-1 \text{ fois}}, \quad \text{puis acceptation.}$$

Par indépendance,

$$\mathbb{P}(N = k) = (1 - p_{\text{acc}})^{k-1} p_{\text{acc}}, \qquad k \geq 1.$$

Donc

$$N \sim \text{Géométrique}\left(\frac{1}{c}\right), \qquad \mathbb{E}(N) = c.$$

Exemple : demi-normale par rejet

On veut simuler

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2} \mathbf{1}_{\{x \geq 0\}},$$

la loi de $|Z|$ lorsque $Z \sim \mathcal{N}(0,1)$. On propose une exponentielle

$$g_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}.$$

Alors, pour $x \geq 0$,

$$\frac{f(x)}{g_\lambda(x)} = \frac{1}{\lambda} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2} + \lambda x \right).$$

En complétant le carré,

$$-\frac{x^2}{2} + \lambda x = -\frac{1}{2}(x - \lambda)^2 + \frac{\lambda^2}{2}.$$

Le maximum est atteint en $x = \lambda$, donc

$$c_\lambda = \frac{1}{\lambda} \sqrt{\frac{2}{\pi}} e^{\lambda^2/2}.$$

Demi-normale : choix de l'enveloppe

Pour minimiser le coût moyen, on minimise c_λ. Il est équivalent de minimiser

$$\log c_\lambda = -\log \lambda + \frac{\lambda^2}{2} + \text{constante.}$$

Sa dérivée vaut

$$-\frac{1}{\lambda} + \lambda,$$

qui s'annule en $\lambda = 1$.

Meilleur choix dans cette famille

$$\lambda = 1, \qquad c_1 = \sqrt{\frac{2e}{\pi}} \approx 1.315.$$

La probabilité d'acceptation est

$$\frac{1}{c_1} \approx 0.760.$$

Idée à retenir

La méthode est efficace : on accepte environ 76% des propositions.

Simulation uniforme dans le disque : méthode de rejet

On simule (X, Y) uniforme sur le carré $[-1, 1]^2$ et on accepte lorsque

$$X^2 + Y^2 \leq 1.$$

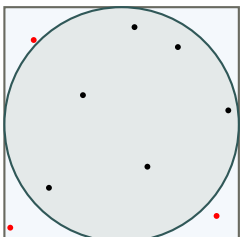
La probabilité d'acceptation vaut

$$\frac{\text{Aire}(B(0,1))}{\text{Aire}([-1,1]^2)} = \frac{\pi}{4}.$$

Pour tout borélien $A \subset B(0,1)$,

$$\mathbb{P}((X,Y) \in A \mid X^2 + Y^2 \leq 1) = \frac{\text{Aire}(A)/4}{\pi/4} = \frac{\text{Aire}(A)}{\pi}.$$

La loi conditionnelle est uniforme sur le disque.



noir : accepté ; rouge : rejeté

Simulation uniforme dans le disque : coordonnées polaires

Écrivons un point du disque sous la forme

$$(R\cos\Theta, R\sin\Theta).$$

Par symétrie, $\Theta \sim \mathcal{U}([0, 2\pi])$. Pour le rayon,

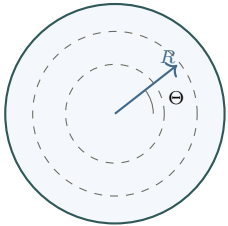
$$\mathbb{P}(R \le r) = \frac{\text{Aire}(B(0,r))}{\text{Aire}(B(0,1))} = r^2, \qquad 0 \le r \le 1.$$

Ainsi la densité de R est

$$f_R(r) = 2r\,\mathbf{1}_{[0,1]}(r).$$

Par inversion,

$$R = \sqrt{U}, \qquad U \sim \mathcal{U}([0,1]).$$



$\mathbb{P}(R \le r) = r^2$

Point de vigilance

Prendre $R \sim \mathcal{U}([0,1])$ est faux : cela surreprésente le centre. L'aire croît comme r^2, pas comme r.

Synthèse de la séance

Messages à retenir

1. La méthode delta linéarise la propagation d'erreur.
2. Les grilles cartésiennes paient la dimension dans le nombre total de points ; Monte-Carlo garde le taux statistique $N^{-1/2}$.
3. Un IC doit être accompagné de diagnostics : stabilité de la moyenne, de S_n, répétitions indépendantes, queues et valeurs extrêmes.
4. L'inversion donne une simulation exacte lorsque l'inverse généralisée est calculable.
5. L'acceptation–rejet est générale ; sa validité vient d'un calcul de loi conditionnelle, et son coût moyen est c.