Méthodes de Monte Carlo

Master parcours SSD - UE Statistique Computationnelle

Septembre 2019

Définition

- Diverses définitions proposées pour caractériser une méthode de Monte Carlo.
- Nous prendrons celle donnée dans Rizzo (2007, §6.1) : *"toute méthode d'inférence statistique ou d'analyse numérique s'appuyant sur des techniques de simulation [de variables aléatoires]".*
- Nous nous intéresserons donc à ces deux types d'applications :
  - l'analyse numérique et la problème de l'intégration.
  - l'inférence statistique pour la caractérisation d'un estimateur ou des performances d'un test statistique.
- Mais avant cela, nous allons nous intéresser à la simulation de variables aléatoires

Simulation de variables aléatoires

Simulation de variables aléatoires

- Simuler des variables selon une loi uniforme est un problème bien connu.
- Le logiciel R permet de simuler selon les lois usuelles.
  - uniforme, normale, Poisson, χ^2, binomiale, ...
- La simulation d'autres lois peut être plus complexe.
- Nous allons illustrer deux méthodes de simulation :
  - l'inversion de la fonction de répartition,
  - l'algorithme du rejet.

Simulation de lois usuelles avec R

- La base : la loi normale (centrée réduite)
  - `dnorm` : fonction densité ($\texttt{dnorm(0)} = 1/\sqrt{2\pi}$)
  - `pnorm` : fonction de répartition (`pnorm(0) = 0.5`)
  - `qnorm` : quantiles ($\texttt{qnorm(0.5)} = 0$)
  - `rnorm` : génération de nombres aléatoires

Simulation de lois usuelles avec R

- La base : la loi normale (centrée réduite)
  - `dnorm` : fonction densité (`dnorm(0)` $= 1/\sqrt{2\pi}$)
  - `pnorm` : fonction de répartition (`pnorm(0) = 0.5`)
  - `qnorm` : quantiles (`qnorm(0.5)` $= 0$)
  - `rnorm` : génération de nombres aléatoires

- Pour les autres lois, remplacer `norm` par un autre suffixe
  - `dunif, punif, qunif, runif` pour la loi uniforme
  - `dexp, pexp, qexp, rexp` pour la loi exponentielle

Simulation de lois usuelles avec R

- La base : la loi normale (centrée réduite)
  - `dnorm` : fonction densité (`dnorm(0)` $= 1/\sqrt{2\pi}$)
  - `pnorm` : fonction de répartition (`pnorm(0) = 0.5`)
  - `qnorm` : quantiles (`qnorm(0.5)` $= 0$)
  - `rnorm` : génération de nombres aléatoires

- Pour les autres lois, remplacer `norm` par un autre suffixe
  - `dunif, punif, qunif, runif` pour la loi uniforme
  - `dexp, pexp, qexp, rexp` pour la loi exponentielle

- Pour visualiser une distribution empirique
  - la fonction `hist` calcule et/ou affiche l'histogramme
  - la fonction `density` calcule la densité par la méthode des noyaux (`plot.density` pour la visualiser)

`?distributions()` : densités disponibles en R

RDocumentation | Search for packages, functions, etc

Details

The functions for the density/mass function, cumulative distribution function, quantile function and random variate generation are n… `qxxx` and `rxxx` respectively.

For the beta distribution see `dbeta` .

For the binomial (including Bernoulli) distribution see `dbinom` .

For the Cauchy distribution see `dcauchy` .

For the chi-squared distribution see `dchisq` .

For the exponential distribution see `dexp` .

For the F distribution see `df` .

For the gamma distribution see `dgamma` .

For the geometric distribution see `dgeom` . (This is also a special case of the negative binomial.)

For the hypergeometric distribution see `dhyper` .

For the log-normal distribution see `dlnorm` .

For the multinomial distribution see `dmultinom` .

For the negative binomial distribution see `dnbinom` .

For the normal distribution see `dnorm` .

For the Poisson distribution see `dpois` .

For the Student's t distribution see `dt` .

For the uniform distribution see `dunif` .

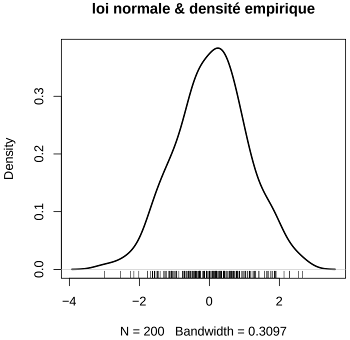
For the Weibull distribution see `dweibull` .

For less common distributions of test statistics see `pbirthday` , `dsignrank` , `ptukey` and `dwilcox` (and see the 'See Also' section of

Simulation de lois usuelles avec R

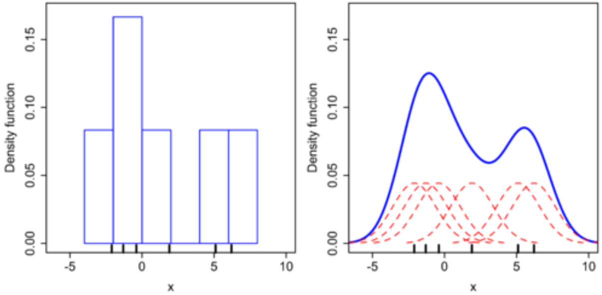
- Exemple :

```
> n = 1000        # nombre d'échantillons
> x = rnorm(n)    # tirage selon la loi N(0,1)
> plot(density(x), main = "")
> title("loi normale & densité empirique"
> rug(x)
```



(Rappel : estimation par noyau - principe)

Principe :



- on positionne un "noyau" sur chaque observation
- on les moyenne pour estimer la densité

$\Rightarrow$ méthode de Parzen : Kernel Density Estimation

(Rappel : estimation par noyau - définition)

Formellement, à partir de l'échantillon $(x_1, ..., x_n)$:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K(x - x_i), \quad \forall x \in \mathcal{X}$$

où $K(.)$ est un noyau = une fonction :

- non-négative
- dont l'intégrale vaut 1
- qui est centrée sur zéro

(Rappel : estimation par noyau - définition)

Formellement, à partir de l'échantillon $(x_1, ..., x_n)$:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K(x - x_i), \quad \forall x \in \mathcal{X}$$

où $K(.)$ est un noyau = une fonction :

- non-négative
- dont l'intégrale vaut 1
- qui est centrée sur zéro

⇒ Intuitivement : une moyenne locale, avec une notion de proximité définie par K.

(Rappel : estimation par noyau - définition)

Formellement, à partir de l'échantillon $(x_1, ..., x_n)$:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K(x - x_i), \quad \forall x \in \mathcal{X}$$

où $K(.)$ est un noyau = une fonction :

- non-négative
- dont l'intégrale vaut 1
- qui est centrée sur zéro

⇒ Intuitivement : une moyenne locale, avec une notion de proximité définie par K.

Noyau typique = Gaussien : $K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$.

(Rappel : estimation par noyau - fonction noyau)

Noyaux classiques :

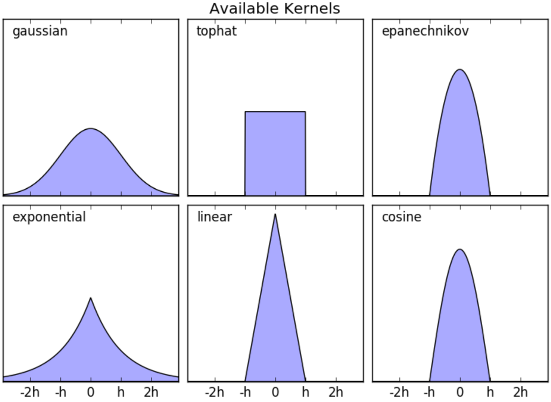


Figure: Noyaux disponibles dans `Scikit-Learn` (et `R`).

(Rappel : estimation par noyau - fonction noyau)

Une question clé : le choix de la largeur de bande

$$\hat{f}(x) = \frac{1}{n}\sum_i K(x - x_i) \Rightarrow \hat{f}_h(x) = \frac{1}{nh}\sum_i K(\frac{x - x_i}{h})$$

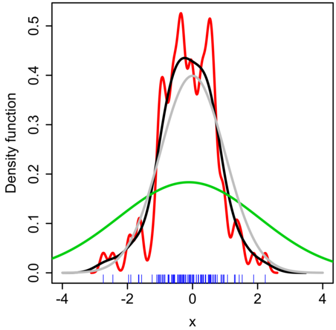


Figure: réalité, h=2, h=0.05, h=0.337

Simulation de variables aléatoires par inversion

Principe :

- S'appuyer sur la distribution cumulée F pour simuler selon f.
- En effet : $\boxed{\text{si } U \to \mathcal{U}(0,1) \text{ alors } F^{-1}(U) \to f}$.

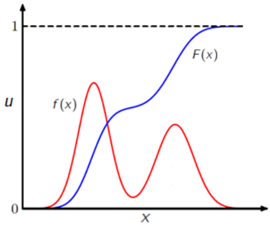
Simulation de variables aléatoires par inversion

Principe :

- S'appuyer sur la distribution cumulée F pour simuler selon f.
- En effet : $\boxed{\text{si } U \to \mathcal{U}(0,1) \text{ alors } F^{-1}(U) \to f}$.

Illustration :

- En rouge = densité cible ; en bleu= distribution cumulée.
- 1) On tire u uniformément sur $[0,1]$.
- 2) On prend x^* tel que $F(x^*) = u$.



$\Rightarrow$ On tire u selon l'axe des ordonnées.

$\Rightarrow$ La probabilité de tirer x est faible dans les zones où $F(x)$ est plate.

Figure: Image tirée de Bishop (2006).

Simulation de variables aléatoires par inversion

Hypothèses de travail :

- on connait la forme analytique de f
- (on sait simuler selon $\mathcal{U}(0,1)$)

Simulation de variables aléatoires par inversion

Hypothèses de travail :

- on connait la forme analytique de f
- (on sait simuler selon $\mathcal{U}(0,1)$)

Procédure :

1. calculer la fonction de repartition $F(x)$
2. calculer sa fonction réciproque $F^{-1}(u)$
   - poser $u = F(x)$
   - résoudre l'équation en x pour trouver $x = F^{-1}(u)$
3. tirer $(u_1, ..., u_n)$ selon $\mathcal{U}(0,1)$
4. calculer $x_i = F^{-1}(u_i)$, pour $i = 1, ..., n$.

Simulation de variables aléatoires par rejet

Principe :

- On choisit 1) une densité auxiliaire g selon laquelle on sait simuler, et 2) $k \in \mathbb{R}$ tel que $\boxed{f(x) \leq kg(x),\ \forall x}$.

Simulation de variables aléatoires par rejet

Principe :

- On choisit 1) une densité auxiliaire g selon laquelle on sait simuler, et 2) $k \in \mathbb{R}$ tel que $\boxed{f(x) \leq kg(x), \ \forall x}$.

Illustration :

- En rouge = densité f ; en bleu= "densité" majorante kg.
- 1) On tire x_0 selon g.
- 2) On tire u_0 uniformément dans $[0; kg(x_0)]$.
- 3) Si $u_0 \leq f(x_0)$ on garde x_0, sinon on le rejette.

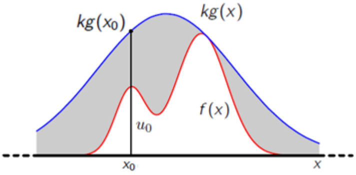


Figure: Image tirée de Bishop (2006).

$\Rightarrow$ La probabilité de tirer x dépend de l'écart entre $f(x)$ et $kg(x)$.

$\Rightarrow$ Le taux de rejet augmente en fonction de l'aire grise.

Simulation de variables aléatoires par rejet

Hypothèses de travail :

- on connait la forme analytique de f
- on connait k et g tels que $f(x) \leq kg(x), \forall x$
- on sait simuler selon g (et selon $\mathcal{U}(0, 1)$)

Simulation de variables aléatoires par rejet

Hypothèses de travail :

- on connait la forme analytique de f
- on connait k et g tels que $f(x) \leq kg(x), \forall x$
- on sait simuler selon g (et selon $\mathcal{U}(0,1)$)

Procédure :

1. tirer x_i selon g, pour $i = 1, ..., n$
2. tirer u_i selon $\mathcal{U}(0, kg(x_i))$
3. conserver x_i si $u_i \leq f(x_i)$

Simulation de variables aléatoires par rejet

Hypothèses de travail :

- on connait la forme analytique de f
- on connait k et g tels que $f(x) \leq kg(x), \forall x$
- on sait simuler selon g (et selon $\mathcal{U}(0,1)$)

Procédure :

1. tirer x_i selon g, pour $i = 1, ..., n$
2. tirer u_i selon $\mathcal{U}(0, kg(x_i))$
3. conserver x_i si $u_i \leq f(x_i)$

En pratique :

- on applique cette procédure jusqu'à obtenir le nombre de tirages voulu (e.g., avec une boucle "tant que").
- le taux de rejet quantifie l'efficacité de la procédure.

Remarques

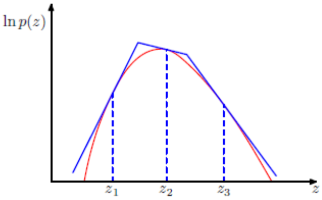
Simulation par inversion :

- $+$: simple
- - : on ne sait pas toujours calculer F^{-1}

Simulation par rejet :

- $+$: plus générique
- - : difficile de choisir la densité majorante

$\Rightarrow$ extension : méthode du rejet adaptatif



$\Rightarrow$ les tirages rejetés servent à définir une enveloppe autour de f.

Figure: Image tirée de Bishop (2006).

Et la loi uniforme dans tout ça ?

La loi uniforme est à la base de nombreux simulateurs.

- via les méthodes d'inversion et de rejet en particulier

Un problème bien connu...mais pas si trivial.

Et la loi uniforme dans tout ça ?

La loi uniforme est à la base de nombreux simulateurs.

- via les méthodes d'inversion et de rejet en particulier

Un problème bien connu...mais pas si trivial.

Méthode classique = générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \quad \text{avec} \quad z_n = (az_{n-1} + b)(\text{modulo } m)}$$

Et la loi uniforme dans tout ça ?

La loi uniforme est à la base de nombreux simulateurs.

- via les méthodes d'inversion et de rejet en particulier

Un problème bien connu...mais pas si trivial.

Méthode classique = générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \quad \text{avec} \quad z_n = (az_{n-1} + b)(\text{modulo } m)}$$

$\Rightarrow$ génère une suite de nombres aléatoires x_i.

Et la loi uniforme dans tout ça ?

La loi uniforme est à la base de nombreux simulateurs.

- via les méthodes d'inversion et de rejet en particulier

Un problème bien connu...mais pas si trivial.

Méthode classique = générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \quad \text{avec} \quad z_n = (az_{n-1} + b)(\text{modulo } m)}$$

$\Rightarrow$ génère une suite de nombres aléatoires x_i.

$\Rightarrow$ à (a, b, m) fixés, la suite est determinée par z_0.

- z_0 est la **graine** (seed) du générateur.

Et la loi uniforme dans tout ça ?

La loi uniforme est à la base de nombreux simulateurs.

- via les méthodes d'inversion et de rejet en particulier

Un problème bien connu...mais pas si trivial.

Méthode classique = générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \quad \text{avec} \quad z_n = (az_{n-1} + b)(\text{modulo } m)}$$

$\Rightarrow$ génère une suite de nombres aléatoires x_i.

$\Rightarrow$ à (a, b, m) fixés, la suite est determinée par z_0.

- z_0 est la **graine** (seed) du générateur.

$\Rightarrow$ c'est en réalité une suite de nombres pseudo-aléatoires.

- on peut donc la répéter en fixant la graine.

Simulation de la loi Uniforme

Méthode du générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \;\; \text{avec} \;\; z_n = (az_{n-1} + b)(\text{modulo } m)}$$

$\Rightarrow$ génère une suite de nombres pseudo-aléatoires

Simulation de la loi Uniforme

Méthode du générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \ \text{ avec } \ z_n = (az_{n-1} + b)(\text{modulo } m)}$$

$\Rightarrow$ génère une suite de nombres pseudo-aléatoires

- z_0 = graine (fixée) ; x_n : n-ième valeur obtenue.
- $z = y(\text{modulo } m)$: le reste de $y/m \rightarrow \in [0, ..., m-1]$
- $x = z/m$: ramène z entre 0 et 1.
- $m \sim$ le nombre de valeurs distinctes possibles.
  - à prendre le + grand possible (e.g., $2^{31} - 1$, 10^8).
- a, b : à choisir avec soin pour avoir une bonne suite !

Simulation de la loi Uniforme

Méthode du générateur congruentiel :

$$\boxed{x_n = \frac{z_n}{m}, \text{ avec } z_n = (az_{n-1} + b)(\text{modulo } m)}$$

⇒ génère une suite de nombres pseudo-aléatoires

- z_0 = graine (fixée) ; x_n : n-ième valeur obtenue.
- $z = y(\text{modulo } m)$: le reste de $y/m \to \in [0, ..., m-1]$
- $x = z/m$: ramène z entre 0 et 1.
- $m \sim$ le nombre de valeurs distinctes possibles.
  - à prendre le + grand possible (e.g., $2^{31} - 1$, 10^8).
- a, b : à choisir avec soin pour avoir une bonne suite !

⇒ voir `?RNG` pour la mise en oeuvre `R`.

⇒ en pratique, utiliser `set.seed()` pour fixer la graine.

- et donc garantir que le script est reproductible.

Méthodes Monte-Carlo pour l'intégration

Motivation

- La question de l'intégration est au cœur de nombreux domaines : physique, finance, biologie...et statistiques.
  - voir 2ème partie du cours sur les approches Bayésiennes
- Parfois complexe à résoudre :
  - nombreuses variables couplées par des modèles complexes
  - primitives difficiles à déterminer
  - primitives trop longues à résoudre par des techniques d'analyse numérique
- L'approche MC s'appuie sur des méthodes de simulation de variables aléatoires pour approximer une intégrale.

Motivation

- La question de l'intégration est au cœur de nombreux domaines : physique, finance, biologie...et statistiques.
  - voir 2ème partie du cours sur les approches Bayésiennes
- Parfois complexe à résoudre :
  - nombreuses variables couplées par des modèles complexes
  - primitives difficiles à déterminer
  - primitives trop longues à résoudre par des techniques d'analyse numérique
- L'approche MC s'appuie sur des méthodes de simulation de variables aléatoires pour approximer une intégrale.

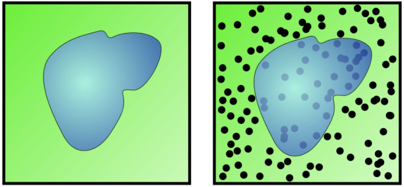
$\Rightarrow$ une approche stochastique pour un problème déterministe.

$\Rightarrow$ approximation = réponse statistique du type *"la valeur recherchée se trouve très probablement dans cet intervalle"*.

Exemples introductifs [1]

Approximation de la superficie d'un lac :

- une armée tire X boulets de canon sur un terrain de taille S.
- On compte ensuite le nombre N de boulets restés sur le terrain.

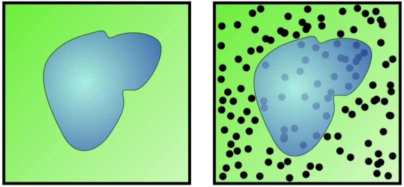


1. Source : Wikipedia

Exemples introductifs [1]

Approximation de la superficie d'un lac :

- une armée tire X boulets de canon sur un terrain de taille S.
- On compte ensuite le nombre N de boulets restés sur le terrain.



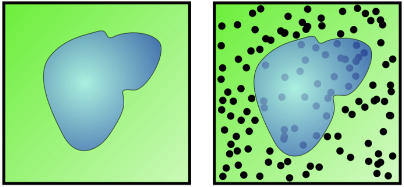
$\Rightarrow$ l'aire du lac peut être approximée comme $S \times \frac{X-N}{X}$.

1. Source : Wikipedia

Exemples introductifs [1]

Approximation de la superficie d'un lac :

- une armée tire X boulets de canon sur un terrain de taille S.
- On compte ensuite le nombre N de boulets restés sur le terrain.



$\Rightarrow$ l'aire du lac peut être approximée comme $S \times \frac{X-N}{X}$.

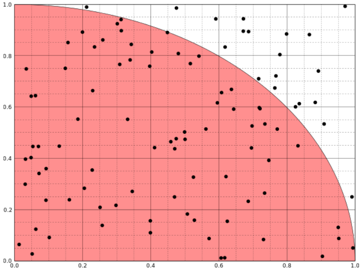
$\Rightarrow$ sous quelle(s) hypothèse(s) est-ce valide ?

1. Source : Wikipedia

Exemples introductifs [2]

Approximation de π :

- on tire aléatoirement (et uniformément) des points (x, y) dans $[0, 1] \times [0, 1]$.
- La proportion de points tels que $x^2 + y^2 \leq 1$ est une approximation de $\pi/4$.



2. Source : Wikipedia

Description de la méthode

- On cherche à calculer

$$I = \int_0^1 g(x)dx.$$

- Principe Monte-Carlo : écrire I comme une espérance.

Description de la méthode

- On cherche à calculer
$$I = \int_0^1 g(x)dx.$$
- Principe Monte-Carlo : écrire I comme une espérance.
- Rappellons que si X est une variable aléatoire de densité f, alors par définition :
$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx.$$

Description de la méthode

- On cherche à calculer

$$I = \int_0^1 g(x)dx.$$

- Principe Monte-Carlo : écrire I comme une espérance.

- Rappellons que si X est une variable aléatoire de densité f, alors par définition :

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx.$$

- Par ailleurs, pour toute fonction $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ est une variable aléatoire d'espérance :

$$E\left[g(X)\right] = \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

Description de la méthode

- On cherche donc à calculer

$$I = \int_0^1 g(x)dx,$$

en l'écrivant comme

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx,$$

où f est une densité de probabilité.

Description de la méthode

- On cherche donc à calculer

$$I = \int_0^1 g(x)dx,$$

en l'écrivant comme

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx,$$

où f est une densité de probabilité.

- Il suffit de considérer que X suit une loi uniforme sur $[0,1]$, sa densité étant définie comme :

$$f(x) = \begin{cases} 1 & \text{si} \quad 0 \leq x \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

Description de la méthode

- On cherche donc à calculer

$$I = \int_0^1 g(x)dx$$

- On l'écrit comme $I = E[g(X)]$: l'espérance de la variable aléatoire $g(X)$, où $X \mapsto \mathcal{U}(0,1)$.

Description de la méthode

- On cherche donc à calculer

$$I = \int_0^1 g(x)dx$$

- On l'écrit comme $I = E[g(X)]$: l'espérance de la variable aléatoire $g(X)$, où $X \mapsto \mathcal{U}(0,1)$.

- Par conséquent, si on dispose d'un n-échantillon $\left(X_1, ..., X_n\right)$ iid de loi $\mathcal{U}(0,1)$, on peut approximer I par l'estimateur de la moyenne empirique :

$$S_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Description de la méthode

- On cherche donc à calculer

$$I = \int_0^1 g(x)dx$$

- On l'écrit comme $I = E[g(X)]$: l'espérance de la variable aléatoire $g(X)$, où $X \mapsto \mathcal{U}(0,1)$.

- Par conséquent, si on dispose d'un n-échantillon $(X_1, ..., X_n)$ iid de loi $\mathcal{U}(0,1)$, on peut approximer I par l'estimateur de la moyenne empirique :

$$S_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

$\Rightarrow$ il suffit de savoir tirer des nombres aléatoires uniformément sur $[0,1]$, i.e., simuler une v.a. uniforme.

Description de la méthode

- En pratique, on s'intéresse souvent à

$$I = \int g(x)f(x)dx,$$

où f est une densité de probabilité quelconque.

Description de la méthode

- En pratique, on s'intéresse souvent à

$$I = \int g(x)f(x)dx,$$

où f est une densité de probabilité quelconque.

- On conserve la forme générale de l'espérance et on interprète I comme

$$I = E\left[g(X)\right],$$

où <u>X est distribuée selon f</u>.

Description de la méthode

- En pratique, on s'intéresse souvent à

$$I = \int g(x)f(x)dx,$$

où f est une densité de probabilité quelconque.

- On conserve la forme générale de l'espérance et on interprète I comme

$$I = E\left[g(X)\right],$$

où X est distribuée selon f.

- On applique le même principe en simulant une variable aléatoire de loi f.

Justification de la méthode (1/2)

Deux théorèmes bien connus permettent de justifier la validité de cette méthode :

Justification de la méthode (1/2)

Deux théorèmes bien connus permettent de justifier la validité de cette méthode :

1. La loi forte des grands nombres qui nous dit que $\bar{X}_n$ converge vers $E(X)$:

$$E(X) = \lim_{n \to +\infty} \bar{X}_n = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- ce résultat nous dit donc que l'approximation est valide.
- (NB : il faut néanmoins que $E(|X|)$ soit intégrable.)

Justification de la méthode (2/2)

Deux théorèmes bien connus permettent de justifier la validité de cette méthode :

2. Le Théorème de la Limite Centrale qui nous dit que

$$\frac{\sqrt{n}}{\sigma}\epsilon_n \to \mathcal{N}(0,1),$$

où $\epsilon_n = E(X) - \bar{X}_n$ est l'erreur d'approximation, et $\sigma^2 = var(X)$.

Justification de la méthode (2/2)

Deux théorèmes bien connus permettent de justifier la validité de cette méthode :

2. Le Théorème de la Limite Centrale qui nous dit que

$$\frac{\sqrt{n}}{\sigma}\epsilon_n \to \mathcal{N}(0,1),$$

   où $\epsilon_n = E(X) - \bar{X}_n$ est l'erreur d'approximation, et $\sigma^2 = var(X)$.

- ce résultat quantifie la vitesse de convergence de notre estimateur :

$$\epsilon_n \to \mathcal{N}(0, \sigma/\sqrt{n}).$$

  - "il converge en racine de n" : il faut 4 fois plus de réalisations pour réduire l'erreur de moitié.

Justification de la méthode (2/2)

Deux théorèmes bien connus permettent de justifier la validité de cette méthode :

2. Le Théorème de la Limite Centrale qui nous dit que

$$\frac{\sqrt{n}}{\sigma}\epsilon_n \to \mathcal{N}(0, 1),$$

où $\epsilon_n = E(X) - \bar{X}_n$ est l'erreur d'approximation, et $\sigma^2 = var(X)$.

- ce résultat quantifie la vitesse de convergence de notre estimateur :

$$\epsilon_n \to \mathcal{N}(0, \sigma/\sqrt{n}).$$

  - "il converge en racine de n" : il faut 4 fois plus de réalisations pour réduire l'erreur de moitié.
- par contre il ne permet pas de borner l'erreur...

Utilisation pratique

- le TLC ne nous permet pas de borner l'erreur, mais il nous permet de donner un intervalle de confiance :
  - Si $N \to \mathcal{N}(0,1)$ alors $p(|N| \leq 1.96) = 0.95$.[3]
  - On a donc $p(|\epsilon_n| < 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$.

3. plus généralement : $p(|N| \leq t_{\alpha/2}) = 1 - \alpha$, où $t_{\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0,1)$.

Utilisation pratique

- le TLC ne nous permet pas de borner l'erreur, mais il nous permet de donner un intervalle de confiance :
  - Si $N \to \mathcal{N}(0,1)$ alors $p(|N| \leq 1.96) = 0.95.$[3]
  - On a donc $p(|\epsilon_n| < 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$.

$\Rightarrow$ l'intervalle de confiance à 95% de $E(X)$ est donc :

$$\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}\ ;\ \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

3. plus généralement : $p(|N| \leq t_{\alpha/2}) = 1 - \alpha$, où $t_{\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0,1)$.

Utilisation pratique

- le TLC ne nous permet pas de borner l'erreur, mais il nous permet de donner un intervalle de confiance :
  - Si $N \to \mathcal{N}(0,1)$ alors $p(|N| \leq 1.96) = 0.95.$ [3]
  - On a donc $p(|\epsilon_n| < 1.96\frac{\sigma}{\sqrt{n}}) = 0.95.$

$\Rightarrow$ l'intervalle de confiance à 95% de $E(X)$ est donc :

$$\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \; ; \; \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

- En pratique, on ne connait pas la variance théorique σ^2 et on l'estime par la variance empirique :

$$\bar{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

3. plus généralement : $p(|N| \leq t_{\alpha/2}) = 1 - \alpha$, où $t_{\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0,1)$.

En résumé

On cherche à calculer $I = \int g(x)f(x)dx$, où f est une densité de probabilité :

En résumé

On cherche à calculer $I = \int g(x)f(x)dx$, où f est une densité de probabilité :

1. On simule un n-échantillon $(X_1, \ldots, X_n)$ selon la loi f.

En résumé

On cherche à calculer $I = \int g(x)f(x)dx$, où f est une densité de probabilité :

1. On simule un n-échantillon $(X_1, \dots, X_n)$ selon la loi f.
2. On calcule :

$$S_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \quad \text{et} \quad V_n = \frac{1}{n-1}\sum_{i=1}^{n}(g(X_i) - S_n)^2.$$

En résumé

On cherche à calculer $I = \int g(x)f(x)dx$, où f est une densité de probabilité :

1. On simule un n-échantillon $(X_1, \ldots, X_n)$ selon la loi f.
2. On calcule :

$$S_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \quad \text{et} \quad V_n = \frac{1}{n-1}\sum_{i=1}^{n}(g(X_i) - S_n)^2.$$

3. On donne un intervalle de confiance sur I défini comme :

$$\left[S_n - t_{\alpha/2}\sqrt{V_n/n}\,;\ S_n + t_{\alpha/2}\sqrt{V_n/n}\right],$$

   où $t_{\alpha/2}$ est le quantile d'ordre $(1-\alpha/2)$ de la loi $\mathcal{N}(0,1)$, pour un intervalle de confiance à $1-\alpha$.

   - (en général on prend $\alpha = 0.05$ et $t_{\alpha/2} = 1.96$ pour un intervalle de confiance à 95%).

Remarques

- Cette méthode est simple à mettre en œuvre.
  - Le seul pré-requis est de savoir simuler des variables aléatoires selon une loi d'intérêt.

Remarques

- Cette méthode est simple à mettre en œuvre.
  - Le seul pré-requis est de savoir simuler des variables aléatoires selon une loi d'intérêt.
- Sa précision augmente (i.e., la largeur de l'intervalle de confiance décroit) en fonction de $\sqrt{n}$, quelle que soit la dimension du problème.
  - faible dimension : relativement lent par rapport aux méthodes déterministes.
  - haute dimension : parfois la seule approche donnant une solution dans un temps raisonnable.

Remarques

- Cette méthode est simple à mettre en œuvre.
  - Le seul pré-requis est de savoir simuler des variables aléatoires selon une loi d'intérêt.
- Sa précision augmente (i.e., la largeur de l'intervalle de confiance décroit) en fonction de $\sqrt{n}$, quelle que soit la dimension du problème.
  - faible dimension : relativement lent par rapport aux méthodes déterministes.
  - haute dimension : parfois la seule approche donnant une solution dans un temps raisonnable.
- En pratique, elle peut être gourmande en temps de calcul à cause (1) de sa faible vitesse de convergence et (2) du coût calculatoire de g qui peut être élevé.
  - les méthodes de réduction de variance permettent d'accélérer la vitesse de convergence de l'algorithme.

Exemple 1

- On veut calculer $I = \int_0^1 e^{-x} dx$.
- La solution est $I = 1 - e^{-1} = 0.6321$.

4. en bloquant la graine du générateur de nombres aléatoires à 20.

Exemple 1

- On veut calculer $I = \int_0^1 e^{-x} dx$.
- La solution est $I = 1 - e^{-1} = 0.6321$.

- On peut l'approximer en R par :

```
> n = 1000        # nombre de tirages
> x = runif(n)    # tirage selon la loi uniforme
> gx = exp(-x)
> I.hat = mean(gx)
```

ce qui donne[4] 0.6307344.

4. en bloquant la graine du générateur de nombres aléatoires à 20.

Exemple 1

- On veut calculer $I = \int_0^1 e^{-x} dx$.
- La solution est $I = 1 - e^{-1} = 0.6321$.

- On peut l'approximer en R par :

```
> n = 1000       # nombre de tirages
> x = runif(n)   # tirage selon la loi uniforme
> gx = exp(-x)
> I.hat = mean(gx)
```

ce qui donne[4] 0.6307344.

- On peut également donner un intervalle de confiance :

```
> alpha = 0.05
> a = qnorm(1-(alpha/2))
> I1 = I.hat - a*sqrt(var(gx)/n)
> I2 = I.hat + a*sqrt(var(gx)/n)
```

ce qui donne [0.6193152; 0.6421535].

4. en bloquant la graine du générateur de nombres aléatoires à 20.

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Première approche : se baser sur $\mathcal{U}(0,1)$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Première approche : se baser sur $\mathcal{U}(0,1)$

- Problème : il faut ramener les limites de l'intégrale à $[0,1]$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Première approche : se baser sur $\mathcal{U}(0,1)$

- Problème : il faut ramener les limites de l'intégrale à $[0,1]$
- Solution = changement de variable :

$$\text{prendre } y = \frac{x-a}{b-a} \text{ soit } \begin{cases} x = (b-a)y + a \\ dx = (b-a)dy \end{cases}$$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Première approche : se baser sur $\mathcal{U}(0,1)$

- Problème : il faut ramener les limites de l'intégrale à $[0,1]$
- Solution = changement de variable :

$$\text{prendre } y = \frac{x-a}{b-a} \text{ soit } \begin{cases} x = (b-a)y + a \\ dx = (b-a)dy \end{cases}$$

- On a donc :

$$I = (b-a)\int_0^1 \exp\Big((a-b)y - a\Big)dy$$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Première approche : se baser sur $\mathcal{U}(0,1)$

- Problème : il faut ramener les limites de l'intégrale à $[0,1]$
- Solution = changement de variable :

$$\text{prendre } y = \frac{x-a}{b-a} \quad \text{soit} \quad \begin{cases} x = (b-a)y + a \\ dx = (b-a)dy \end{cases}$$

- On a donc :

$$I = (b-a) \int_0^1 \exp\Big((a-b)y - a\Big) dy$$

- Exemple :

```
> m = 1000; a = 2; b = 4
> y = runif(m)   # tirage selon la loi uniforme(0,1)
> Ihat.1 = (b-a)*mean( exp((a-b)*y-a) )
```

ce qui donne 0.1187561 (au lieu de 0.1170196).

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Deuxième approche : tirer dans $\mathcal{U}(a, b)$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Deuxième approche : tirer dans $\mathcal{U}(a, b)$

- Problème :
  - la fonction $f(x) = \mathbf{1}(x \in [a, b])$ n'est pas une densité

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Deuxième approche : tirer dans $\mathcal{U}(a, b)$

- Problème :
  - la fonction $f(x) = \mathbf{1}(x \in [a, b])$ n'est pas une densité
  - la densité de la loi $\mathcal{U}(a, b)$ est $f(x) = \frac{1}{b-a}\mathbf{1}(x \in [a, b])$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Deuxième approche : tirer dans $\mathcal{U}(a, b)$

- Problème :
  - la fonction $f(x) = \mathbf{1}(x \in [a, b])$ n'est pas une densité
  - la densité de la loi $\mathcal{U}(a, b)$ est $f(x) = \frac{1}{b-a}\mathbf{1}(x \in [a, b])$
- Solution = faire apparaître la densité $\mathcal{U}(a, b)$:

$$
\begin{aligned}
I &= \int_a^b e^{-x} dx \\
&= (b-a) \int_a^b e^{-x} \frac{1}{b-a} dx
\end{aligned}
$$

Exemple 2 : on veut calculer $I = \int_a^b e^{-x} dx$

$\Rightarrow$ Deuxième approche : tirer dans $\mathcal{U}(a, b)$

- Problème :
  - la fonction $f(x) = \mathbf{1}(x \in [a, b])$ n'est pas une densité
  - la densité de la loi $\mathcal{U}(a, b)$ est $f(x) = \frac{1}{b-a}\mathbf{1}(x \in [a, b])$
- Solution = faire apparaître la densité $\mathcal{U}(a, b)$:

$$
\begin{aligned}
I &= \int_a^b e^{-x} dx \\
&= (b-a) \int_a^b e^{-x} \frac{1}{b-a} dx
\end{aligned}
$$

- Exemple :

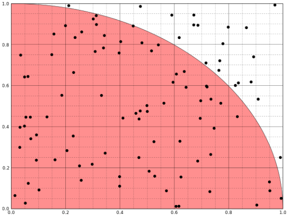
```
> z = runif(m, a, b)
> Ihat.2 = (b-a) * mean( exp(-z) )
```

ce qui donne 0.1147875.

Quizz

Revenons à notre exemple introductif :

- on tire aléatoirement des points (x, y) dans $[0, 1] \times [0, 1]$.
- on approxime $\pi/4$ par la proportion de points tels que $x^2 + y^2 \leq 1$.

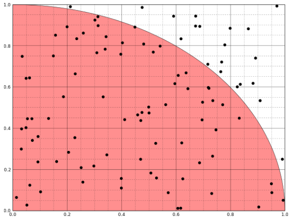


Comment peut-on l'écrire formellement sous la forme d'un problème Monte-Carlo ?

Quizz

Revenons à notre exemple introductif :

- on tire aléatoirement des points (x, y) dans $[0, 1] \times [0, 1]$.
- on approxime $\pi/4$ par la proportion de points tels que $x^2 + y^2 \leq 1$.



Comment peut-on l'écrire formellement sous la forme d'un problème Monte-Carlo ?

$$\Rightarrow \text{celui d'approximer } I = \int_0^1 \int_0^1 \mathbf{1}\big(x^2 + y^2 \leq 1\big)\, dx dy.$$

Remarques & conclusions

- Simulation de variables aléatoires :
  - méthodes par inversion et par rejet
  - place centrale de la loi $\mathcal{U}(0,1)$
  - pour aller plus loin : rejet adaptatif et échantillonnage préférentiel.

Remarques & conclusions

- Simulation de variables aléatoires :
  - méthodes par inversion et par rejet
  - place centrale de la loi $\mathcal{U}(0,1)$
  - pour aller plus loin : rejet adaptatif et échantillonnage préférentiel.
- Méthodes MC pour l'intégration :
  - approche stochastique à un problème déterministe
    - solution = estimation + intervalle de confiance
  - parfois la seule solution envisageable
    - e.g., en physique et en finance.
  - attention aux domaines de définition de l'intégrale et de la densité à simuler.
    - changement de variable, normalisation de la densité
  - pour aller plus loin : méthodes de réduction de variance.

Pour aller plus loin...
méthodes de réduction de variance

Méthodes de réduction de variance

Objectif :

- Améliorer la vitesse de convergence de l'estimateur d'intégrale / d'espérance.
- L'estimateur MC standard fait une erreur $\epsilon \to N(0, \frac{\sigma}{\sqrt{n}})$: on cherche donc à diminuer sa variance (à n fixé).

Méthodes de réduction de variance

Objectif :

- Améliorer la vitesse de convergence de l'estimateur d'intégrale / d'espérance.
- L'estimateur MC standard fait une erreur $\epsilon \to N(0, \frac{\sigma}{\sqrt{n}})$: on cherche donc à diminuer sa variance (à n fixé).

Principe :

- trouver des moyens de ré-écrire $I = E[g(X)]$ comme $I = E[h(Y)]$, tels que $var(h(Y)) \leq var(g(X))$.

Méthodes de réduction de variance

Objectif :

- Améliorer la vitesse de convergence de l'estimateur d'intégrale / d'espérance.
- L'estimateur MC standard fait une erreur $\epsilon \to N(0, \frac{\sigma}{\sqrt{n}})$: on cherche donc à diminuer sa variance (à n fixé).

Principe :

- trouver des moyens de ré-écrire $I = E[g(X)]$ comme $I = E[h(Y)]$, tels que $var(h(Y)) \leq var(g(X))$.

Plusieurs approches :

- échantillonnage préférentiel ("importance sampling"),
- utilisation de variables antithétiques,
- utilisation de variables de contrôle,
- (stratification).

Echantillonnage préférentiel

Principe :

- Introduire une nouvelle densité $\tilde{f}$, et ré-écrire l'intégrale :

$$I = \int g(x)f(x)dx = \int \frac{g(x)f(x)}{\tilde{f}(x)}\tilde{f}(x)dx.$$

Echantillonnage préférentiel

Principe :

- Introduire une nouvelle densité $\tilde{f}$, et ré-écrire l'intégrale :

$$I = \int g(x)f(x)dx = \int \frac{g(x)f(x)}{\tilde{f}(x)}\tilde{f}(x)dx.$$

- On a donc $E[g(X)] = E[\frac{g(Y)f(Y)}{\tilde{f}(Y)}]$, où X est distribuée selon f et Y est distribuée selon $\tilde{f}$.

$\Rightarrow$ Nouveau schéma avantageux si $var(\frac{g(Y)f(Y)}{\tilde{f}(Y)}) < var(g(X))$.

Echantillonnage préférentiel

Principe :

- Introduire une nouvelle densité $\tilde{f}$, et ré-écrire l'intégrale :

$$I = \int g(x)f(x)dx = \int \frac{g(x)f(x)}{\tilde{f}(x)}\tilde{f}(x)dx.$$

- On a donc $E[g(X)] = E[\frac{g(Y)f(Y)}{\tilde{f}(Y)}]$, où X est distribuée selon f et Y est distribuée selon $\tilde{f}$.

$\Rightarrow$ Nouveau schéma avantageux si $var(\frac{g(Y)f(Y)}{\tilde{f}(Y)}) < var(g(X))$.

En pratique :

- Il faut choisir une densité $\tilde{f}$ proche de $|g \times f|$.
- Il faut savoir selon simuler selon $\tilde{f}$.
- $\tilde{f}$ s'appelle la fonction d'importance.

Réduction de variance par variables antithétiques

Principe :

- Pour approximer $I = \int_0^1 g(x)dx$ utiliser le fait que si $U \to \mathcal{U}(0,1)$, alors $(1-U) \to \mathcal{U}(0,1)$.

Réduction de variance par variables antithétiques

Principe :

- Pour approximer $I = \int_0^1 g(x)dx$ utiliser le fait que si $U \to \mathcal{U}(0,1)$, alors $(1-U) \to \mathcal{U}(0,1)$.
- On peut donc estimer $I = E\left[g(U)\right]$ à n fixé par :

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \quad \text{ou} \quad \frac{1}{n}\sum_{i=1}^{n/2}\Big(g(X_i) + g(1-X_i)\Big).$$

Réduction de variance par variables antithétiques

Principe :

- Pour approximer $I = \int_0^1 g(x)dx$ utiliser le fait que si $U \to \mathcal{U}(0,1)$, alors $(1-U) \to \mathcal{U}(0,1)$.
- On peut donc estimer $I = E\left[g(U)\right]$ à n fixé par :

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \quad \text{ou} \quad \frac{1}{n}\sum_{i=1}^{n/2}\Big(g(X_i) + g(1-X_i)\Big).$$

$\Rightarrow$ Nouveau schéma toujours plus efficace si g est monotone, car $g(U)$ et $g(1-U)$ sont alors anti-corrélées.

- et $var(A+B) = var(A) + var(B) + 2cov(A,B)$

Réduction de variance par variables antithétiques

Principe :

- Pour approximer $I = \int_0^1 g(x)dx$ utiliser le fait que si $U \to \mathcal{U}(0,1)$, alors $(1-U) \to \mathcal{U}(0,1)$.
- On peut donc estimer $I = E\big[g(U)\big]$ à n fixé par :

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \quad \text{ou} \quad \frac{1}{n}\sum_{i=1}^{n/2}\Big(g(X_i) + g(1-X_i)\Big).$$

$\Rightarrow$ Nouveau schéma toujours plus efficace si g est monotone, car $g(U)$ et $g(1-U)$ sont alors anti-corrélées.

- et $var(A+B) = var(A) + var(B) + 2cov(A,B)$

En pratique :

- n'est valable que si g est continue et monotone.
- U et $(1-U)$ sont dites antithétiques.

Réduction de variance par variables de contrôle

Principe :

- Pour approximer $I = \int g(x)dx$, introduire une fonction h proche de g qui soit facilement intégrable.
- On peut alors écrire :

$$I = E\left[g(X)\right] = E\left[g(X) - h(X)\right] + E\left[h(X)\right]$$

$\Rightarrow$ Nouveau schéma avantageux si $var\left(g(X) - h(X)\right) < var\left(g(X)\right)$

Réduction de variance par variables de contrôle

Principe :

- Pour approximer $I = \int g(x)dx$, introduire une fonction h proche de g qui soit facilement intégrable.
- On peut alors écrire :

$$I = E\left[g(X)\right] = E\left[g(X) - h(X)\right] + E\left[h(X)\right]$$

$\Rightarrow$ Nouveau schéma avantageux si $var\left(g(X) - h(X)\right) < var\left(g(X)\right)$

En pratique :

- Il faut donc trouver h qui soit proche de g et que l'on sache intégrer (*i.e.*, que l'on sache calculer $E\left[h(X)\right]$).
- Le fait que h et g soient corrélées devrait garantir que $var\left(g(X) - h(X)\right)$ soit faible.
- $h(X)$ est la variable de contrôle de $g(X)$.

Références

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

Maria L. Rizzo. *Statistical Computing with R*. CRC Press, 2007.