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 (`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

Simulation de lois usuelles avec R

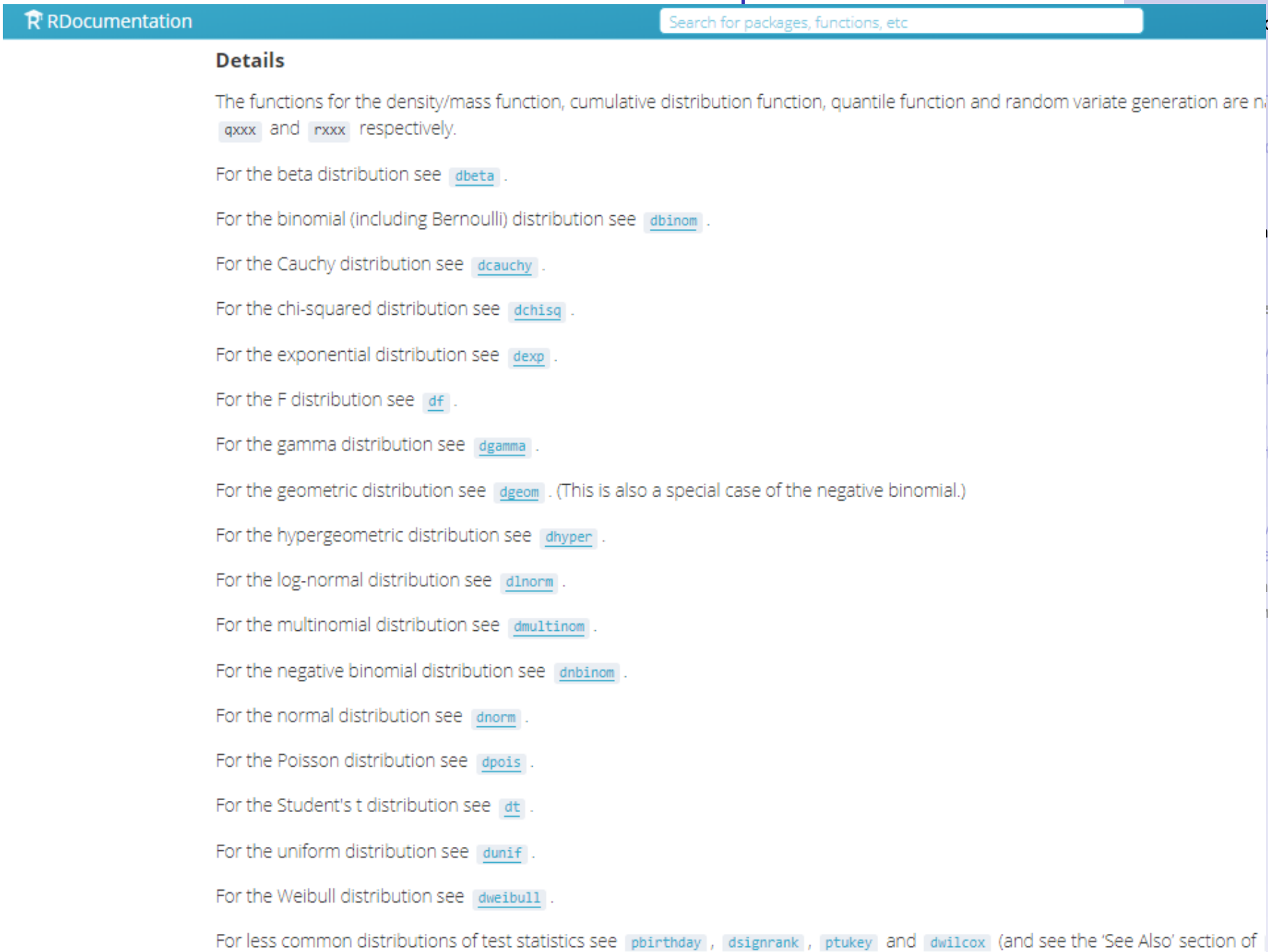
- La base : la loi normale (centrée réduite)
  - `dnorm` : fonction densité (`dnorm(0)` $= 1/\sqrt{2\pi}$)
  - `pnorm` : fonction de répartition (`pnorm(0) = 0.5`)
  - `qnorm` : quantiles (`qnorm(0.5)` $= 0$)
  - `rnorm` : génération de nombres aléatoires

- 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` .

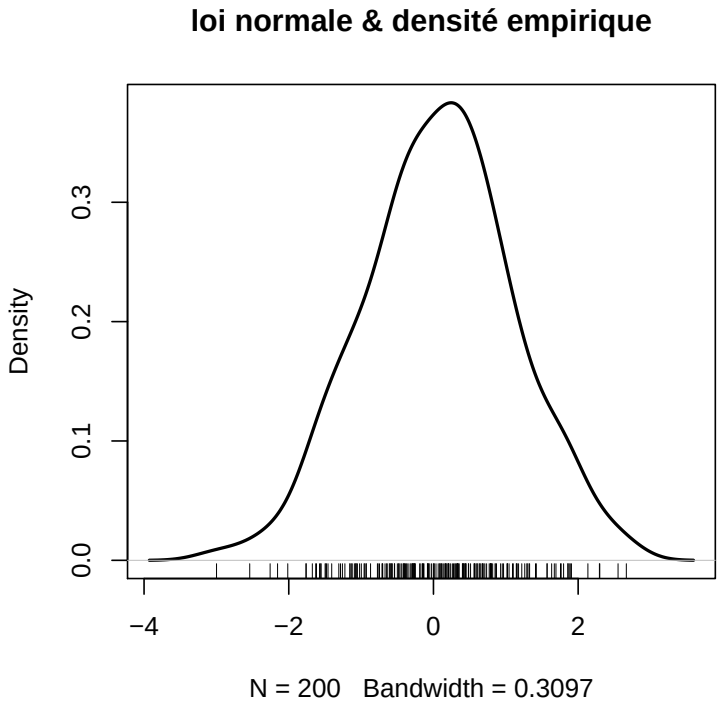
For the Weibull distribution see `dweibull` .

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

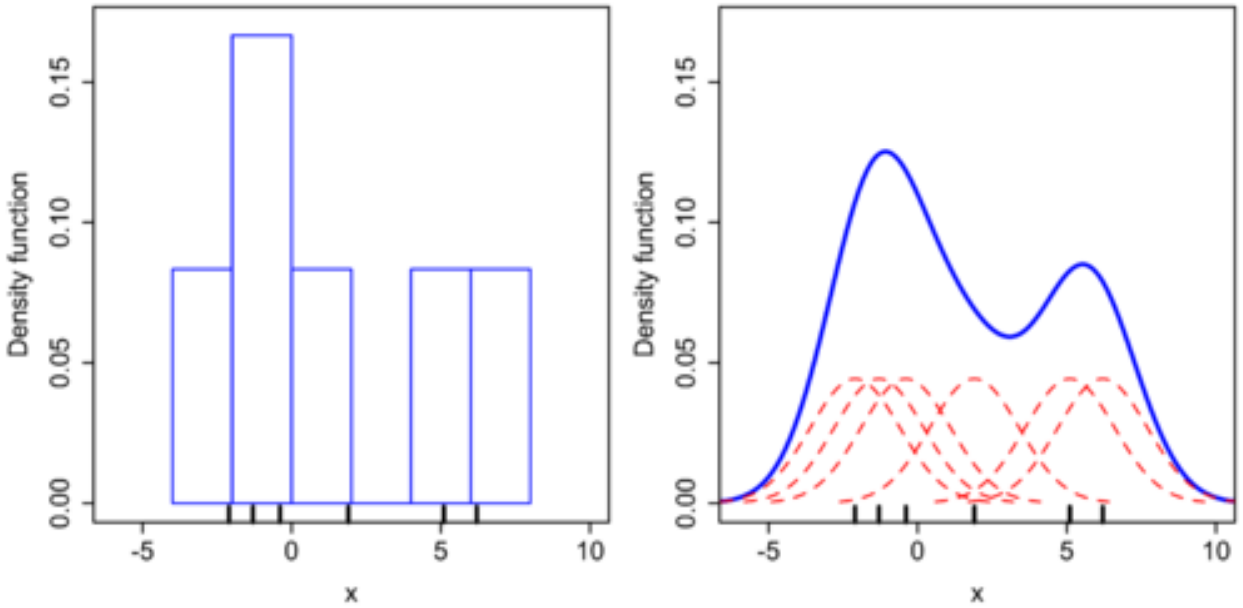
- Exemple :

```
> 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

$\Rightarrow$ 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

$\Rightarrow$ 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 :

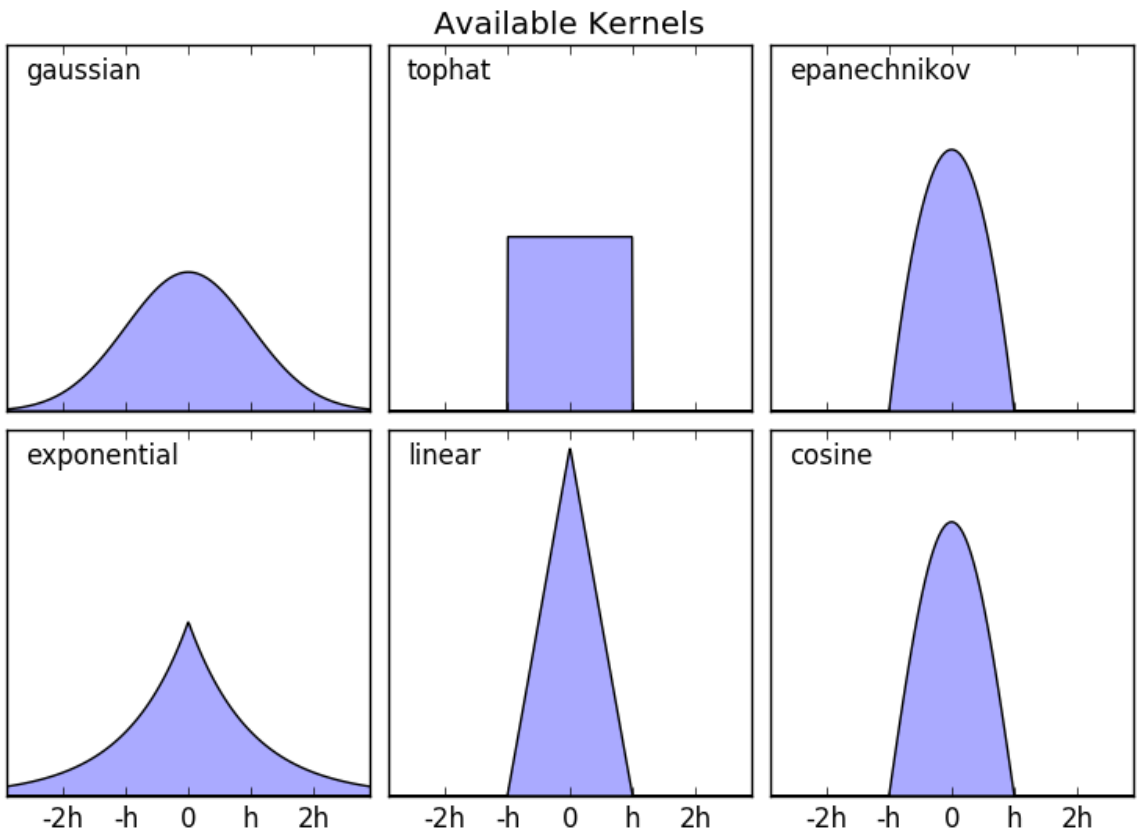


Figure: Noyaux disponibles dans `Scikit-Learn` (et `R`).

(Rappel : estimation par noyau - fonction noyau)

Une question clé : le choix de la largeur de bande

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

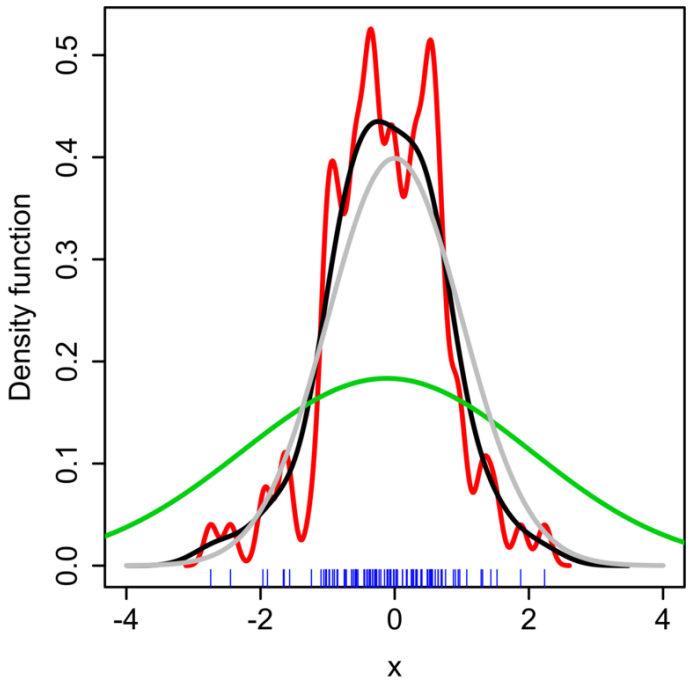


Figure: réalité, h=2, h=0.05, h=0.337

Simulation de variables aléatoires par inversion

Principe :

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

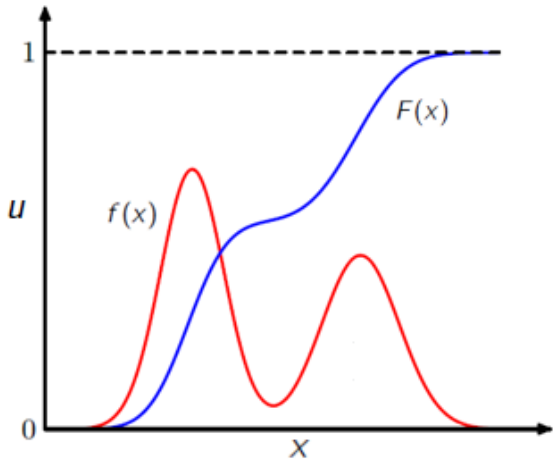
Simulation de variables aléatoires par inversion

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.

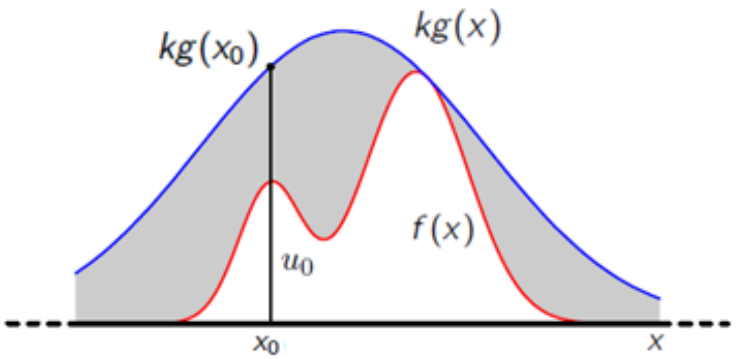


Figure: Image tirée de Bishop (2006).

$\Rightarrow$ 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

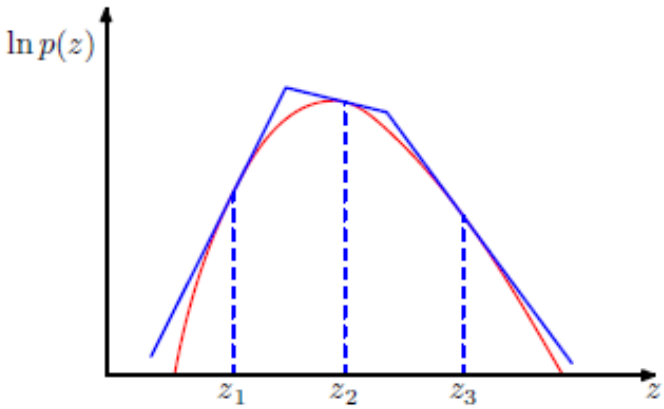
Simulation par inversion :

- + : 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 :

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

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

$\Rightarrow$ à (a, b, m) fixés, la suite est déterminée 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 :

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

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

$\Rightarrow$ à (a, b, m) fixés, la suite est déterminée 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 :

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

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

Simulation de la loi Uniforme

Méthode du générateur congruentiel :

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

$\Rightarrow$ 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 :

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

$\Rightarrow$ 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 !

$\Rightarrow$ voir `?RNG` pour la mise en oeuvre `R`.

$\Rightarrow$ 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.

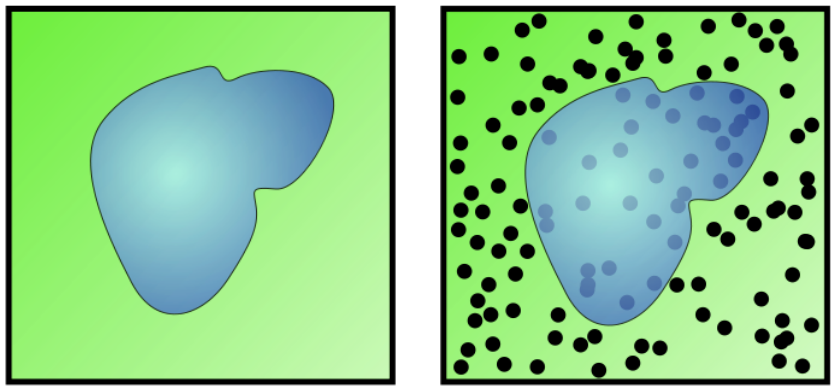
$\Rightarrow$ une approche stochastique pour un problème déterministe.

$\Rightarrow$ 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.

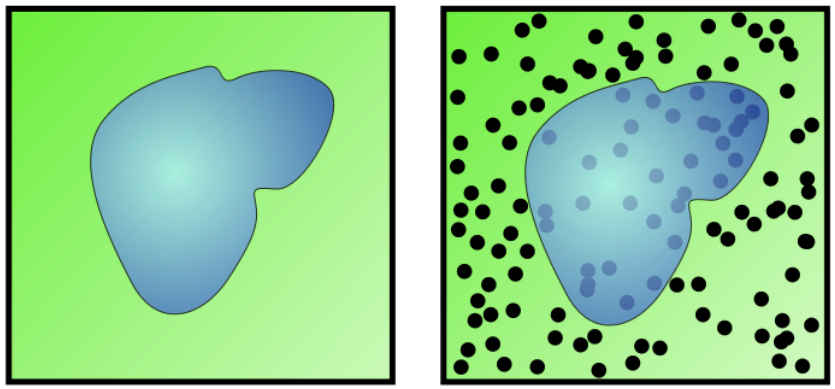


1. Source : Wikipedia

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.



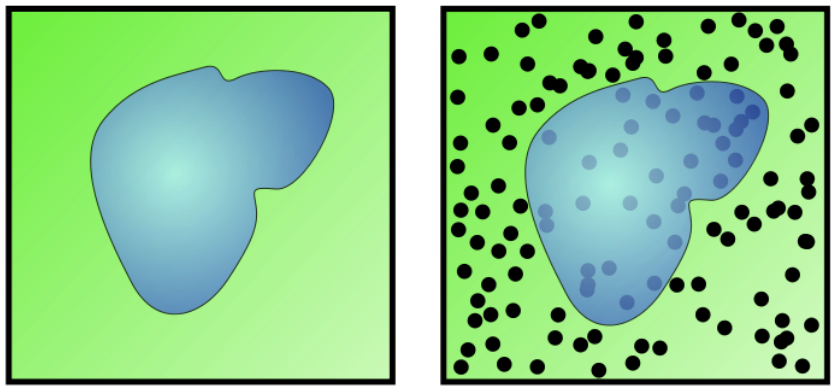
$\Rightarrow$ l'aire du lac peut être approximée comme $S \times \frac{X-N}{X}$.

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 $S \times \frac{X-N}{X}$.

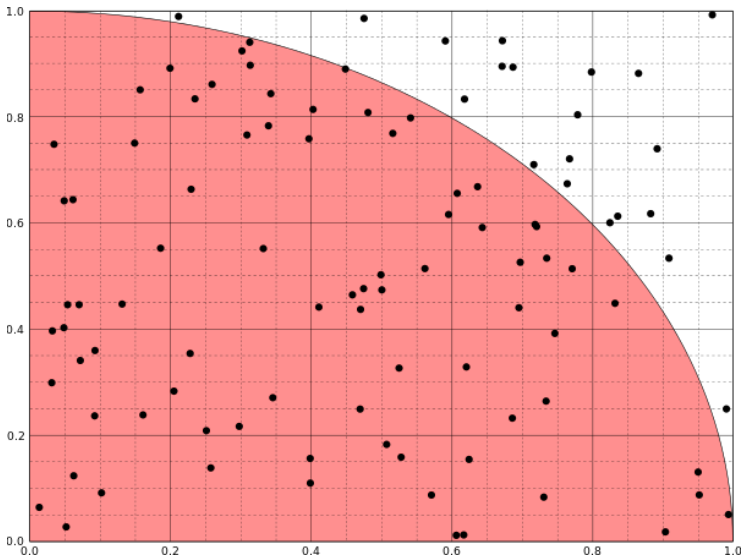
$\Rightarrow$ sous quelle(s) hypothèse(s) est-ce valide ?

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 $(X_1, ..., X_n)$ iid de loi $\mathcal{U}(0,1)$, on peut approximer I par l'estimateur de la moyenne empirique :

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

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

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

où X est distribuée selon f.

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

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

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 $(X_1, \ldots, X_n)$ selon la loi f.
2. On calcule :

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

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 :

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

- On a donc :

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

- Exemple :

```
> 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 $\mathcal{U}(a, b)$:

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

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 $\mathcal{U}(a, b)$:

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

- Exemple :

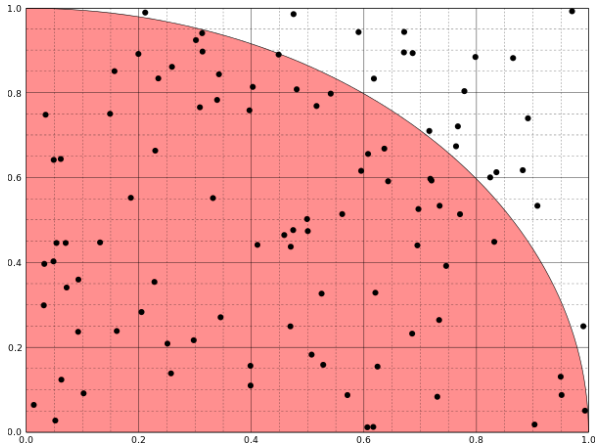
```
> z = runif(m, a, b)
> Ihat.2 = (b-a) * mean( exp(-z) )
```

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 $x^2 + y^2 \leq 1$.

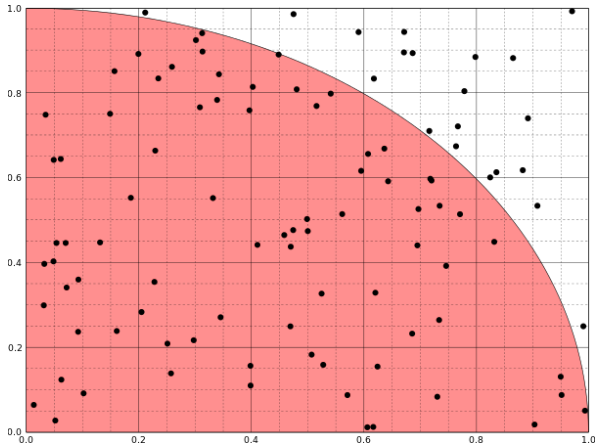


Comment peut-on l'écrire formellement sous la forme d'un problème Monte-Carlo ?

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 ?

$$\Rightarrow \text{celui d'approximer } I = \int_0^1 \int_0^1 \mathbf{1}\left(x^2 + y^2 \leq 1\right) dxdy.$$

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 $I = E\big[g(U)\big]$ à n fixé par :

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

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

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 :

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

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

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 :

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

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

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 $E\big[h(X)\big]$).
- Le fait que h et g soient corrélées devrait garantir que $var\big(g(X) - h(X)\big)$ soit faible.
- $h(X)$ est la variable de contrôle de $g(X)$.

Méthodes Monte-Carlo pour l'inférence statistique

Introduction

Inférence statistique :

- Induire les caractéristiques d'une population à partir d'un échantillon (issu de cette population).
- Deux grandes questions : fournir des estimations de ces caractéristiques et faire des tests d'hypothèses.

Introduction

Inférence statistique :

- Induire les caractéristiques d'une population à partir d'un échantillon (issu de cette population).
- Deux grandes questions : fournir des estimations de ces caractéristiques et faire des tests d'hypothèses.

Méthodes Monte-Carlo pour l'inférence :

- Tirer des échantillons à partir d'un modèle probabiliste de la population.
- Evaluer empiriquement l'incertitude de l'estimation.

Introduction

Inférence statistique :

- Induire les caractéristiques d'une population à partir d'un échantillon (issu de cette population).
- Deux grandes questions : fournir des estimations de ces caractéristiques et faire des tests d'hypothèses.

Méthodes Monte-Carlo pour l'inférence :

- Tirer des échantillons à partir d'un modèle probabiliste de la population.
- Evaluer empiriquement l'incertitude de l'estimation.

$\Rightarrow$ Applications :

- étudier la distribution d'échantillonnage d'un estimateur
- estimer les propriétés d'un test statistique

Estimateur / Estimation

Soit $(X_1, ..., X_n)$ un n-échantillon distribué selon la loi de X.

- Un estimateur $\hat{\theta}$ d'un paramètre θ est une fonction de l'échantillon :

$$\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$$

Estimateur / Estimation

Soit $(X_1, ..., X_n)$ un n-échantillon distribué selon la loi de X.

- Un estimateur $\hat{\theta}$ d'un paramètre θ est une fonction de l'échantillon :
$$\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$$
- C'est lui même une variable aléatoire qui possède sa propre distribution.
- On parle de distribution d'échantillonnage (sampling distribution).

Estimateur / Estimation

Soit $(X_1, ..., X_n)$ un n-échantillon distribué selon la loi de X.

- Un estimateur $\hat{\theta}$ d'un paramètre θ est une fonction de l'échantillon :
$$\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$$
- C'est lui même une variable aléatoire qui possède sa propre distribution.
- On parle de distribution d'échantillonnage (sampling distribution).

- Une estimation est la valeur de l'estimateur pour une réalisation $(x_1, ..., x_n)$ de l'échantillon.

Méthodes MC & inférence - caractérisation d'un estimateur et intervalles de confiance

Méthodes MC pour l'estimation

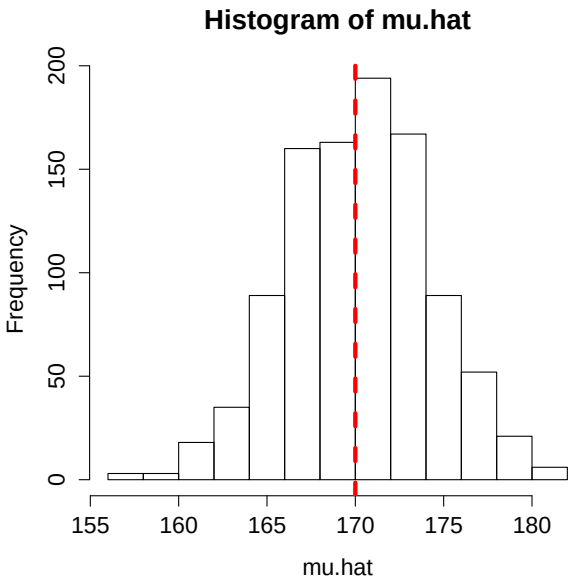
- L'approche MC (couverte ici) vise à caractériser les propriétés d'un estimateur d'une grandeur que l'on connaît (et donc qu'on peut contrôler / fixer).
- Typiquement : le paramètre d'une loi de probabilité
  - (on parle parfois de bootstrap paramétrique)

Méthodes MC pour l'estimation

- L'approche MC (couverte ici) vise à caractériser les propriétés d'un estimateur d'une grandeur que l'on connaît (et donc qu'on peut contrôler / fixer).
- Typiquement : le paramètre d'une loi de probabilité
  - (on parle parfois de bootstrap paramétrique)
- Elle consiste à :
  1. tirer m n-échantillons $\left(X_1^{(j)}, ..., X_n^{(j)}\right)_{j=1,...,m}$, en fixant le paramètre θ à estimer.
  2. calculer les m estimations $\hat{\theta}^{(j)}$, $j = 1, ..., m$.
  3. étudier la distribution d'échantillonnage de $\hat{\theta}$ à partir de ces m réalisations.

Illustration

- On fait l'hypothèse que la taille des étudiants est distribuée normalement selon $\mathcal{N}\big(\mu = 170, \sigma = 20\big)$.
- Illustration de la distribution d'échantillonnage de l'estimateur "moyenne empirique" de μ : variabilité attendue sur 1000 échantillons de $n = 25$ élèves.




```
> m = 1000; n  = 25
> mu  = 170; sigma = 20
> mu.hat  = replicate(m,
expr = {x=rnorm(n,mu,sigma); mean(x)}))
> hist(mu.hat)
> abline(v=mu, lty=2, lwd=2, col=2)
```

Motivations : à quoi ça sert ?

- Les paramètres des lois usuelles sont bien connus.
  - on dispose d'estimateurs performants (e.g., non biaisés et de variance minimale).
  - on connait leur distribution d'échantillonnage : on peut leur associer des intervalles de confiance.

Motivations : à quoi ça sert ?

- Les paramètres des lois usuelles sont bien connus.
  - on dispose d'estimateurs performants (e.g., non biaisés et de variance minimale).
  - on connait leur distribution d'échantillonnage : on peut leur associer des intervalles de confiance.
- Leurs propriétés sont souvent basées sur des hypothèses (e.g., de normalité) et/ou des résulats asymptotiques.

Motivations : à quoi ça sert ?

- Les paramètres des lois usuelles sont bien connus.
  - on dispose d'estimateurs performants (e.g., non biaisés et de variance minimale).
  - on connait leur distribution d'échantillonnage : on peut leur associer des intervalles de confiance.
- Leurs propriétés sont souvent basées sur des hypothèses (e.g., de normalité) et/ou des résulats asymptotiques.
- Dans les applications réelles, ces hypothèses ne sont pas toujours vérifiées
  - pas toujours tout à fait normal, peu d'observations.

Motivations : à quoi ça sert ?

- Les paramètres des lois usuelles sont bien connus.
  - on dispose d'estimateurs performants (e.g., non biaisés et de variance minimale).
  - on connait leur distribution d'échantillonnage : on peut leur associer des intervalles de confiance.
- Leurs propriétés sont souvent basées sur des hypothèses (e.g., de normalité) et/ou des résulats asymptotiques.
- Dans les applications réelles, ces hypothèses ne sont pas toujours vérifiées
  - pas toujours tout à fait normal, peu d'observations.

$\Rightarrow$ l'approche MC permet (entre autres) de quantifier l'impact d'hypothèses non vérifiées sur les propriétés de l'estimateur.

Caractérisation d'un estimateur

Pour caractériser un estimateur $\hat{\theta}$, on peut s'intéresser :

- à son biais : $\text{Biais}(\hat{\theta}) = E\left[\hat{\theta}\right] - \theta$
- à son erreur quadratique moyenne : $\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$
- à son erreur type $\text{se}(\hat{\theta})$, définie comme l'écart type de sa distribution d'échantillonnage.

Caractérisation d'un estimateur

Pour caractériser un estimateur $\hat{\theta}$, on peut s'intéresser :

- à son biais : $\text{Biais}(\hat{\theta}) = E\left[\hat{\theta}\right] - \theta$
- à son erreur quadratique moyenne : $\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$
- à son erreur type $\text{se}(\hat{\theta})$, définie comme l'écart type de sa distribution d'échantillonnage.

Ces critères permettent notamment :

- de caractériser la précision d'un estimateur en fonction de la taille n de l'échantillon
- de comparer la performance de différents estimateurs

Illustration : estimer la moyenne d'une loi normale

On souhaite estimer la moyenne d'une loi normale à partir d'un échantillon de taille $n = 20$.

Illustration : estimer la moyenne d'une loi normale

On souhaite estimer la moyenne d'une loi normale à partir d'un échantillon de taille $n = 20$.

Estimateur naturel : moyenne empirique : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- On sait qu'il est sans biais (loi des grands nombres)
- On connait son erreur-type : $\text{se}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$.
  - $var(X_1 + ... + X_n) = n\sigma^2$, donc $var(\bar{X}_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

Illustration : estimer la moyenne d'une loi normale

On souhaite estimer la moyenne d'une loi normale à partir d'un échantillon de taille $n = 20$.

Estimateur naturel : moyenne empirique : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- On sait qu'il est sans biais (loi des grands nombres)
- On connait son erreur-type : $\text{se}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$.
  - $var(X_1 + ... + X_n) = n\sigma^2$, donc $var(\bar{X}_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$
- En pratique, on ne connait pas σ^2 et on utilise la variance empirique comme estimateur de la variance :

$$\hat{se}(\bar{x_n}) = \frac{1}{\sqrt{n}}\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}.$$

Illustration : estimer la moyenne d'une loi normale

On souhaite estimer la moyenne d'une loi normale à partir d'un échantillon de taille $n = 20$.

Estimateur naturel : moyenne empirique : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- On sait qu'il est sans biais (loi des grands nombres)
- On connait son erreur-type : $\text{se}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$.
  - $var(X_1 + ... + X_n) = n\sigma^2$, donc $var(\bar{X}_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$
- En pratique, on ne connait pas σ^2 et on utilise la variance empirique comme estimateur de la variance :

$$\hat{se}(\bar{x_n}) = \frac{1}{\sqrt{n}}\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}.$$

$\Rightarrow$ pourquoi aller chercher plus loin ?

Illustration : estimer la moyenne d'une loi normale

Et si nos observations n'étaient pas tout à fait normales mais "contaminées" par 1% d'"outliers" :

$$X \to 0.99\mathcal{N}(0,1) + 0.01\mathcal{N}(0,10)$$

Illustration : estimer la moyenne d'une loi normale

Et si nos observations n'étaient pas tout à fait normales mais "contaminées" par 1% d'"outliers" :

$$X \to 0.99\mathcal{N}(0,1) + 0.01\mathcal{N}(0,10)$$

D'autres estimateurs pourraient être plus robustes :

Illustration : estimer la moyenne d'une loi normale

Et si nos observations n'étaient pas tout à fait normales mais "contaminées" par 1% d'"outliers" :

$$X \to 0.99\mathcal{N}(0,1) + 0.01\mathcal{N}(0,10)$$

D'autres estimateurs pourraient être plus robustes :

- la médiane,

Illustration : estimer la moyenne d'une loi normale

Et si nos observations n'étaient pas tout à fait normales mais "contaminées" par 1% d'"outliers" :

$$X \to 0.99\mathcal{N}(0,1) + 0.01\mathcal{N}(0,10)$$

D'autres estimateurs pourraient être plus robustes :

- la médiane,
- la moyenne empirique "trimmée" (trimmed) où on élimine la plus grande et la plus petite observation,

Illustration : estimer la moyenne d'une loi normale

Et si nos observations n'étaient pas tout à fait normales mais "contaminées" par 1% d'"outliers" :

$$X \to 0.99\mathcal{N}(0,1) + 0.01\mathcal{N}(0,10)$$

D'autres estimateurs pourraient être plus robustes :

- la médiane,
- la moyenne empirique "trimmée" (trimmed) où on élimine la plus grande et la plus petite observation,
- la moyenne empirique "trimée" d'ordre k où on supprime les k plus petites et k plus grandes observations.

Illustration : estimer la moyenne d'une loi normale

Et si nos observations n'étaient pas tout à fait normales mais "contaminées" par 1% d'"outliers" :

$$X \to 0.99\mathcal{N}(0,1) + 0.01\mathcal{N}(0,10)$$

D'autres estimateurs pourraient être plus robustes :

- la médiane,
- la moyenne empirique "trimmée" (trimmed) où on élimine la plus grande et la plus petite observation,
- la moyenne empirique "trimée" d'ordre k où on supprime les k plus petites et k plus grandes observations.

$\Rightarrow$ Problème : on ne connaît pas leurs propriétés.

Illustration : estimer la moyenne d'une loi normale

Stratégie MC : simuler des échantillons et caractériser empiriquement ces estimateurs :

Illustration : estimer la moyenne d'une loi normale

Stratégie MC : simuler des échantillons et caractériser empiriquement ces estimateurs :

1. tirer m n-échantillons $\left(X_1^{(j)}, ..., X_n^{(j)}\right)_{j=1,...,m}$, en fixant le paramètre θ à estimer.

Illustration : estimer la moyenne d'une loi normale

Stratégie MC : simuler des échantillons et caractériser empiriquement ces estimateurs :

1. tirer m n-échantillons $\left(X_1^{(j)}, ..., X_n^{(j)}\right)_{j=1,...,m}$, en fixant le paramètre θ à estimer.
2. calculer les m estimations $\hat{\theta}^{(j)}$, $j = 1, ..., m$.

Illustration : estimer la moyenne d'une loi normale

Stratégie MC : simuler des échantillons et caractériser empiriquement ces estimateurs :

1. tirer m n-échantillons $\left(X_1^{(j)}, ..., X_n^{(j)}\right)_{j=1,...,m}$, en fixant le paramètre θ à estimer.
2. calculer les m estimations $\hat{\theta}^{(j)}$, $j = 1, ..., m$.
3. étudier la distribution d'échantillonnage des $\hat{\theta}^{(j)}$:

$$\text{Biais}(\hat{\theta}) : \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}^{(j)} - \theta = \bar{\hat{\theta}} - \theta$$

$$\text{MSE}(\hat{\theta}) : \frac{1}{m}\sum_{j=1}^{m}(\hat{\theta}^{(j)} - \theta)^2$$

$$\text{Erreur type - se}(\hat{\theta}) : \left(\frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{\theta}^{(j)} - \bar{\hat{\theta}}\right)^2\right)^{1/2}$$

(NB : ça n'a en général pas d'importance d'utiliser la version biaisée ou non de l'écart type car on simule en général de nombreux échantillons)

Illustration : estimer la moyenne d'une loi normale - procédure R

```
> n = 20; m = 1000; k = 5
> e = matrix(0, m, 4)
> for(i in 1:m){
x = sort(rnorm(n))
e[i,1] = mean(x)
e[i,2] = mean(x[2:(n-1)])
e[i,3] = mean(x[(k+1):(n-k)])
e[i,4] = median(x)
}
> mse = apply(e, 2, function(x){mean(x^2)})
> se = apply(e, 2, function(x){sqrt(sum((x - mean(x))^2)/m)})

> mse
[1] 0.05165281 0.05317642 0.06306673 0.07978597

> se
[1] 0.2272608 0.2305817 0.2511308 0.2824580
```

$\Rightarrow$ TP : étude de la robustesse de ces estimateurs.

Estimation d'un niveau de confiance - motivation

- En pratique, une estimation s'accompagne souvent d'un niveau de confiance, formalisé comme un intervalle de confiance.

Estimation d'un niveau de confiance - motivation

- En pratique, une estimation s'accompagne souvent d'un niveau de confiance, formalisé comme un intervalle de confiance.
- Ces intervalles sont souvent obtenus sous des hypothèses de normalité de la population.
  - qui peuvent être justifiée si (on pense que) la loi est effectivement normale.
  - qui sont sinon liées à des approximations asymptotiques (e.g., théorème central limite).

Estimation d'un niveau de confiance - motivation

- En pratique, une estimation s'accompagne souvent d'un niveau de confiance, formalisé comme un intervalle de confiance.
- Ces intervalles sont souvent obtenus sous des hypothèses de normalité de la population.
  - qui peuvent être justifiée si (on pense que) la loi est effectivement normale.
  - qui sont sinon liées à des approximations asymptotiques (e.g., théorème central limite).
- On peut appliquer le même type d'approche pour estimer le vrai niveau de confiance d'une procédure d'estimation quand on s'éloigne des hypothèses de normalité.

Estimation d'un niveau de confiance - principe

Soit X la variable aléatoire étudiée et θ le paramètre à estimer (à partir d'un échantillon de taille n).

On va s'appuyer sur une procédure de Monte Carlo suivante :

- Pour chaque répétition $j = 1, ..., m$:
  - générer le $j$ème n-échantillon $\left(X_1^{(j)}, ..., X_n^{(j)}\right)$.
  - calculer l'intervalle de confiance C_j correspondant.
  - vérifier si $\theta \in C_j$.

Le niveau de confiance empirique est égal à la proportion d'intervalles de confiance contenant θ.

Estimation d'un niveau de confiance - illustration

- On cherche à estimer la variance σ^2 d'une variable aléatoire X.

Estimation d'un niveau de confiance - illustration

- On cherche à estimer la variance σ^2 d'une variable aléatoire X.
- Si elle est normalement distribuée, et qu'on dispose d'un n-échantillon $(X_1, ..., X_n)$, alors

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1),$$

où S_n^2 est la variance empirique.

Estimation d'un niveau de confiance - illustration

- On cherche à estimer la variance σ^2 d'une variable aléatoire X.
- Si elle est normalement distribuée, et qu'on dispose d'un n-échantillon $(X_1, ..., X_n)$, alors

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1),$$

où S_n^2 est la variance empirique.

- Un intervalle de confiance à $100(1-\alpha)\%$ pour σ^2 est donné par :

$$\left[(n-1)S_n^2/\chi^2_{1-\alpha/2} \;;\; (n-1)S_n^2/\chi^2_{\alpha/2}\right],$$

où χ^2_α est le quantile d'ordre α de la distribution $\chi^2(n-1)$.

Estimation d'un niveau de confiance - illustration

- On peut vérifier la définition de cet intervalle de confiance en simulant une loi normale :

```
> m = 1000; n = 20; sigma = 2; alpha = 0.05
> I1 = numeric(m); I2 = numeric(m)
> for(i in 1:1000){
x = rnorm(n, mean = 0, sd = sigma)
I1[i] = (n-1)*var(x)/qchisq(1-alpha/2, df = n-1)
I2[i] = (n-1)*var(x)/qchisq(alpha/2, df = n-1)}
> print( mean(sigma^2 > I1 & sigma^2 < I2) )
```

Estimation d'un niveau de confiance - illustration

- On peut vérifier la définition de cet intervalle de confiance en simulant une loi normale :

```
> m = 1000; n = 20; sigma = 2; alpha = 0.05
> I1 = numeric(m); I2 = numeric(m)
> for(i in 1:1000){
x = rnorm(n, mean = 0, sd = sigma)
I1[i] = (n-1)*var(x)/qchisq(1-alpha/2, df = n-1)
I2[i] = (n-1)*var(x)/qchisq(alpha/2, df = n-1)}
> print( mean(sigma^2 > I1 & sigma^2 < I2) )
```

- Cette procédure nous donne – comme attendu – approximativement 95%.
- On sait néanmoins que cette définition d'intervalle de confiance est assez sensible aux écarts à la normalité...

$\Rightarrow$ TP : évaluer la robustesse de cette procédure.

Méthodes MC et estimation - résumé

En s'appuyant sur des techniques de simulation, l'approche MC permet de caractériser empiriquement la précision d'un estimateur en fonction de la taille de l'échantillon.

Méthodes MC et estimation - résumé

En s'appuyant sur des techniques de simulation, l'approche MC permet de caractériser empiriquement la précision d'un estimateur en fonction de la taille de l'échantillon.

C'est une approche souvent plus simple à mettre en oeuvre que des développements mathématiques visant à affiner les approximations asymptotiques.

Méthodes MC et estimation - résumé

En s'appuyant sur des techniques de simulation, l'approche MC permet de caractériser empiriquement la précision d'un estimateur en fonction de la taille de l'échantillon.

C'est une approche souvent plus simple à mettre en oeuvre que des développements mathématiques visant à affiner les approximations asymptotiques.

Elle permet notamment de comparer la performance de plusieurs estimateurs et d'évaluer empiriquement les niveaux de confiance associés quand on s'éloigne de leurs hypothèses de validité.

- taille d'échantillon et/ou loi de la variable aléatoire.

Méthodes MC & inférence - tests statistiques

Méthodes MC et tests statistiques - introduction

Test d'hypothèse : évaluer la validité d'une hypothèse statistique en fonction d'un échantillon.

- valeur théorique vs estimation et fluctuation d'échantillonnage.

Méthodes MC et tests statistiques - introduction

Test d'hypothèse : évaluer la validité d'une hypothèse statistique en fonction d'un échantillon.

- valeur théorique vs estimation et fluctuation d'échantillonnage.

Faire un choix entre deux hypothèses statistiques :

- l'hypothèse nulle notée H_0
- une hypothèse alternative notée H_1

Méthodes MC et tests statistiques - introduction

Test d'hypothèse : évaluer la validité d'une hypothèse statistique en fonction d'un échantillon.

- valeur théorique vs estimation et fluctuation d'échantillonnage.

Faire un choix entre deux hypothèses statistiques :

- l'hypothèse nulle notée H_0
- une hypothèse alternative notée H_1

Démarche générale :

1. Définir une statistique de test et sa distribution sous H_0.
2. Choisir un seuil de significativité, et en déduire la zone de rejet de H_0.
3. Evaluer la statistique de test sur un échantillon et prendre la décision : rejeter ou accepter H_0.

Méthodes MC et tests statistiques - introduction

Exemple : on veut tester l'hypothèse $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ pour une v.a. X de loi $\mathcal{N}(\mu, \sigma)$, de variance σ^2 connue, à partir d'un échantillon de taille n.

Méthodes MC et tests statistiques - introduction

Exemple : on veut tester l'hypothèse $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ pour une v.a. X de loi $\mathcal{N}(\mu, \sigma)$, de variance σ^2 connue, à partir d'un échantillon de taille n.

Procédure :

1. On va se baser sur la moyenne empirique pour estimer μ.

Méthodes MC et tests statistiques - introduction

Exemple : on veut tester l'hypothèse $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ pour une v.a. X de loi $\mathcal{N}(\mu, \sigma)$, de variance σ^2 connue, à partir d'un échantillon de taille n.

Procédure :

1. On va se baser sur la moyenne empirique pour estimer μ.
2. Sous H_0, on sait que $\bar{X}_n \to \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}})$.

Méthodes MC et tests statistiques - introduction

Exemple : on veut tester l'hypothèse $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ pour une v.a. X de loi $\mathcal{N}(\mu, \sigma)$, de variance σ^2 connue, à partir d'un échantillon de taille n.

Procédure :

1. On va se baser sur la moyenne empirique pour estimer μ.
2. Sous H_0, on sait que $\bar{X}_n \to \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}})$.
3. On déduit notre région de rejet au seuil de significativité α : $T = \mu_0 + t_{1-\alpha} \times \sigma/\sqrt{n}$.

Méthodes MC et tests statistiques - introduction

Exemple : on veut tester l'hypothèse $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ pour une v.a. X de loi $\mathcal{N}(\mu, \sigma)$, de variance σ^2 connue, à partir d'un échantillon de taille n.

Procédure :

1. On va se baser sur la moyenne empirique pour estimer μ.
2. Sous H_0, on sait que $\bar{X}_n \to \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}})$.
3. On déduit notre région de rejet au seuil de significativité α : $T = \mu_0 + t_{1-\alpha} \times \sigma/\sqrt{n}$.
4. Si la réalisation $\bar{x_n}$ est supérieure à T, on rejette H_0.

(voir schéma...)

Méthodes MC et tests statistiques - introduction

Deux types d'erreurs :

Méthodes MC et tests statistiques - introduction

Deux types d'erreurs :

- rejeter H_0 à tort = le risque de première espèce.
  - on le note α.
  - il est **défini a priori** : c'est le seuil de significativité choisi.

Méthodes MC et tests statistiques - introduction

Deux types d'erreurs :

- rejeter H_0 à tort = le risque de première espèce.
  - on le note α.
  - il est **défini a priori** : c'est le seuil de significativité choisi.
- accepter H_0 à tort = le risque de seconde espèce
  - on le note β.
  - il est propre à une **hypothèse alternative spécifique**.

Méthodes MC et tests statistiques - introduction

Deux types d'erreurs :

- rejeter H_0 à tort = le risque de première espèce.
  - on le note α.
  - il est **défini a priori** : c'est le seuil de significativité choisi.
- accepter H_0 à tort = le risque de seconde espèce
  - on le note β.
  - il est propre à une **hypothèse alternative spécifique**.

		Décision	
		H_0 vraie	H_0 fausse
Réalité	H_0 vraie	$1-\alpha$	α
	H_0 fausse	β	$1-\beta$

(voir schéma...)

Méthodes MC et tests statistiques - introduction

Deux notions importantes :

Méthodes MC et tests statistiques - introduction

Deux notions importantes :

- la puissance du test = la probabilité de rejeter H_0 à raison ($\sim$ la probabilité de détecter l'hypothèse alternative).
  - elle vaut par définition $1 - \beta$.

Méthodes MC et tests statistiques - introduction

Deux notions importantes :

- la puissance du test = la probabilité de rejeter H_0 à raison ($\sim$ la probabilité de détecter l'hypothèse alternative).
  - elle vaut par définition $1 - \beta$.
- la p-valeur du test = la probabilité d'observer sous H_0 une valeur plus élevée de la statistique de test que celle observée sur l'échantillon.
  - le plus faible α auquel on aurait pu rejeter l'hypothèse nulle compte tenu de notre observation.

(voir schéma...)

Méthodes MC et tests statistiques

L'approche MC peut être déclinée pour étudier les performances d'un test statistique en terme :

- de risque de première espèce : le risque (empirique) de rejeter à tort l'hypothèse nulle est-il conforme à celui attendu ?
- de puissance : estimer empiriquement la probabilité de rejeter l'hypothèse nulle **pour une hypothèse alternative donnée**.

Méthodes MC et tests statistiques

L'approche MC peut être déclinée pour étudier les performances d'un test statistique en terme :

- de risque de première espèce : le risque (empirique) de rejeter à tort l'hypothèse nulle est-il conforme à celui attendu ?
- de puissance : estimer empiriquement la probabilité de rejeter l'hypothèse nulle **pour une hypothèse alternative donnée**.

Cette approche peut notamment être utile pour évaluer la performance d'un test quand le nombre d'observations est limité, ou pour comparer la puissance de différents tests.

- dimensionnement du "sample size" de l'étude

Méthodes MC et tests statistiques

Procédure pour mesurer empiriquement le risque de 1ère espèce d'un test :

Méthodes MC et tests statistiques

Procédure pour mesurer empiriquement le risque de 1ère espèce d'un test :

- Pour $j = 1, .., m$
  - générer le j-ème n-échantillon $\left(X_1^{(j)}, ..., X_n^{(j)}\right)$ selon l'hypothèse nulle
  - calculer la statistique de test T_j
  - vérifier si l'hypothèse nulle est rejetée ou non (au seuil de significativité voulu)

Méthodes MC et tests statistiques

Procédure pour mesurer empiriquement le risque de 1ère espèce d'un test :

- Pour $j = 1, .., m$
  - générer le j-ème n-échantillon $\left(X_1^{(j)}, ..., X_n^{(j)}\right)$ selon l'hypothèse nulle
  - calculer la statistique de test T_j
  - vérifier si l'hypothèse nulle est rejetée ou non (au seuil de significativité voulu)
- Le risque de 1ère espèce empirique est égal à la proportion de tests rejetés.
  - NB : on les rejette à tort.

Méthodes MC et tests statistiques

Procédure pour mesurer empiriquement la puissance d'un test :

Méthodes MC et tests statistiques

Procédure pour mesurer empiriquement la puissance d'un test :

- Pour $j = 1, .., m$
  - générer le j-ème n-échantillon $\left(X_1^{(j)}, ..., X_n^{(j)}\right)$ selon l'hypothèse alternative à évaluer
  - calculer la statistique de test T_j
  - vérifier si l'hypothèse nulle est rejetée ou non (au seuil de significativité voulu)

Méthodes MC et tests statistiques

Procédure pour mesurer empiriquement la puissance d'un test :

- Pour $j = 1, .., m$
  - générer le j-ème n-échantillon $\left(X_1^{(j)}, ..., X_n^{(j)}\right)$ selon l'hypothèse alternative à évaluer
  - calculer la statistique de test T_j
  - vérifier si l'hypothèse nulle est rejetée ou non (au seuil de significativité voulu)
- La puissance empirique est égale à la proportion de tests rejetés.
  - NB : on les rejette à raison.

Remarques et conclusion

Conclusions

Monte-Carlo pour l'inférence :

- simuler des échantillons à partir d'un modèle probabiliste
- évaluer empiriquement l'incertitude de l'estimation

Conclusions

Monte-Carlo pour l'inférence :

- simuler des échantillons à partir d'un modèle probabiliste
- évaluer empiriquement l'incertitude de l'estimation

4 "recettes" génériques :

1. caractérisation d'un estimateur
2. etimation d'un niveau de confiance
3. estimation du risque de 1ère espèce d'un test
4. estimation de la puissance d'un test

Conclusions

Monte-Carlo pour l'inférence :

- simuler des échantillons à partir d'un modèle probabiliste
- évaluer empiriquement l'incertitude de l'estimation

4 "recettes" génériques :

1. caractérisation d'un estimateur
2. etimation d'un niveau de confiance
3. estimation du risque de 1ère espèce d'un test
4. estimation de la puissance d'un test

$\Rightarrow$ Simple à mettre en oeuvre.

Conclusions

Monte-Carlo pour l'inférence :

- simuler des échantillons à partir d'un modèle probabiliste
- évaluer empiriquement l'incertitude de l'estimation

4 "recettes" génériques :

1. caractérisation d'un estimateur
2. etimation d'un niveau de confiance
3. estimation du risque de 1ère espèce d'un test
4. estimation de la puissance d'un test

$\Rightarrow$ Simple à mettre en oeuvre.

$\Rightarrow$ Utile pour dimensionner un problème et/ou quantifier l'impact de l'écart aux hypothèses.

Remarques

- Méthode Monte Carlo : *toute méthode d'inférence statistique ou d'analyse numérique s'appuyant sur des techniques de simulation [de variables aléatoires]* (Rizzo, 2007, §6.1).
- Une autre application importante non couverte = simulation à large échelle d'un système pour étudier sa sensibilité aux fluctuations de ses entrées.
  - "sensitivity analysis" et/ou "uncertainty analysis"
- L'approche générale décrite dans la section "inférence" est parfois appellée bootstrap paramétrique. Le prochain cours s'intéressera au bootstrap "classique".
  - ré-échantillonnage à partir de l'échantillon.

Mise en oeuvre R : la fonction replicate

Avec les approches MC, on fait beaucoup de boucles...

La fonction `replicate` permet de les faire pour vous :

```
> m = 1000;
> n  = 25
> mu  = 170; sigma = 20
> mu.hat = replicate(m,
            expr = {x = rnorm(n,mu,sigma); mean(x)})
```

Utilisation :

- 1er argument : nombre de réplications à faire
- 2ème argument : calcul à faire
- en sortie : un vecteur contenant les m valeurs obtenues

Références

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

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