

ECN 6338 Cours 8

La génération de variables aléatoires univariées

William McCausland

2023-03-09

Survol du cours 8

1. Suites quasi-aléatoires
2. Suites pseudo-aléatoires uniformes
3. Variables (pseudo)-aléatoires non-uniformes
 - a. Par inversion de la fonction de répartition
 - b. Par la méthode de rejet
 - c. Par la méthode de Ziggurat
4. Variables (pseudo)-aléatoires multivariée “faisables”
 - a. de la loi gaussienne multivariée (généralisation de la loi gaussienne)
 - b. de la loi Dirichlet (généralisation de la loi beta)
 - c. de la loi Wishart (généralisation de la loi gamma)

Suites quasi-aléatoires

1. Suites univariées
2. Suites multivariées

Suites univariées quasi-aléatoires

- Une suite quasi-aléatoire sur l'intervalle $[a, b]$ est une suite $\{x_i\}_{i=1}^{\infty} \subseteq [a, b]$, qui est equidistribuée : pour chaque fonction $f: \mathbb{R} \rightarrow \mathbb{R}$, intégrable dans le sens de Riemann,

$$\lim_{n \rightarrow \infty} \frac{b-a}{n} \sum_{i=1}^n f(x_i) = \int_a^b f(x) dx.$$

Une suite sur $[0, 1]$ qui n'est pas équirépartie, deux autres qui le sont

La suite $\{w_i\}_{i=1}^{\infty}$:

$$w_n = 0, 1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \frac{3}{16}, \dots$$

Elle n'est pas équirépartie, Pour la fonction $f(x) = 1$, la somme suivante ne converge pas :

$$\frac{1}{n} \sum_{i=1}^n f(w_i) = \frac{1}{n} \sum_{i=1}^n w_i.$$

Par contre, la suite $\{x_i\}_{i=1}^{\infty}$ de Halton (pour le nombre premier 2) est semblable, mais équirépartie :

$$x_n = \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \dots$$

Pour θ irrationnel (réel et non une fraction d'entiers), la suite $\{i\theta\}$, $i = 1, \dots$ est équirépartie ($\{x\}$ ici est la partie fractionnaire de x)

Suites univariées pseudo-aléatoires

- ▶ Une suite pseudo-aléatoire est une suite $\{u_i\}_{i=1}^{\infty}$ en \mathbb{R} pour laquelle le modèle $u_i \sim \text{iid } U([0, 1])$ est “bonne”.
- ▶ Souvent on utilise $\{u_i\}_{i=1}^{\infty}$ pour construire une suite $\{X_i\}_{i=1}^{\infty}$, $X_i \in \mathbb{R}^d$, et on démontre que si $u_i \sim \text{iid } U([0, 1])$ alors X_i est iid avec densité cible $q(x)$.
- ▶ Le modèle implique que pour chaque $f: \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $E[f(X)]$ et $\text{Var}[f(X)]$ existent,

$$\hat{I} \equiv \frac{1}{n} \sum_{i=1}^{\infty} f(X_i) \xrightarrow{p.s.} I \equiv E[f(X)] = \int_{-\infty}^{\infty} f(x)q(x) dx,$$

$$\text{Var}[\hat{I}] = \frac{\text{Var}[f(X_i)]}{n}.$$

- ▶ Notez le cas spécial

$$\frac{1}{n} \sum_{i=1}^n f(a + u_i(b - a)) \xrightarrow{p.s.} \frac{1}{b - a} \int_a^b f(x) dx.$$

Suites multivariées quasi-aléatoires

- ▶ Une suite quasi-aléatoire sur $D \subseteq \mathbb{R}^d$ est une suite $\{x_i\}_{i=1}^\infty \subseteq D$ qui est équidistribuée sur D : pour chaque fonction $f: \mathbb{R}^d \rightarrow \mathbb{R}$, intégrable dans le sens de Riemann,

$$\lim_{n \rightarrow \infty} \frac{\mu(D)}{n} \sum_{i=1}^n f(x_i) = \int_D f(x) dx.$$

- ▶ Très souvent, $D = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d]$, auquel cas on peut transformer à $D = [0, 1]^d$.
- ▶ Intuition : on veut des suites univariées “indépendantes”
- ▶ Suite de Weyl (torus, dans le paquet R `quasiRNG`) :

$$(\{i\sqrt{p_1}\}, \{i\sqrt{p_2}\}, \dots, \{i\sqrt{p_d}\}),$$

où p_j est le j -ième nombre premier.

- ▶ La suite de **Sobol** est populaire mais difficile à décrire.
- ▶ La suite de **Halton** est populaire et relativement simple à décrire (diapo suivante)

Illustration de la suite Halton, $d = 2$

i	x_1^i	(binaire)	x_2^i	(base 3)
1	$\frac{1}{2}$	0.1_2	$\frac{1}{3}$	0.1_3
2	$\frac{1}{4}$	0.01_2	$\frac{2}{3}$	0.2_3
3	$\frac{3}{4}$	0.11_2	$\frac{1}{9}$	0.01_3
4	$\frac{1}{8}$	0.001_2	$\frac{4}{9}$	0.11_3
5	$\frac{5}{8}$	0.101_2	$\frac{7}{9}$	0.21_3
6	$\frac{3}{8}$	0.011_2	$\frac{2}{9}$	0.02_3
7	$\frac{7}{8}$	0.111_2	$\frac{5}{9}$	0.12_3
8	$\frac{1}{16}$	0.0001_2	$\frac{8}{9}$	0.22_3
9	$\frac{9}{16}$	0.1001_2	$\frac{1}{27}$	0.001_3

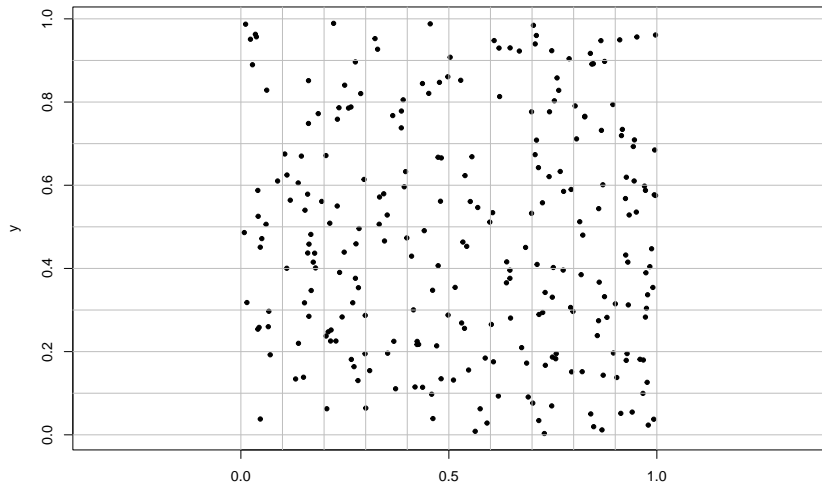
Suites multivariées pseudo-aléatoires

Si $U_i \sim \text{iid } U([0, 1])$, $i = 1, \dots$,

$$V_j = (U_{(j-1)d+1}, U_{(j-1)d+2}, \dots, U_{jd}) \sim \text{iid } U([0, 1]^d)$$

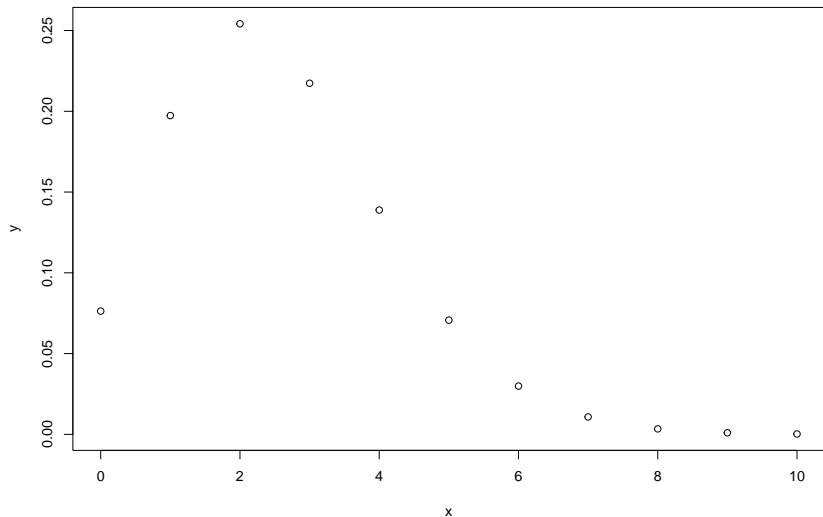
256 points pseudo-aléatoire dans $[0, 1]^2$

```
set.seed(1234567890); x = runif(256); y = runif(256)
plot(x, y, asp=1, pch=20, xlim=c(0,1))
abline(h=seq(0.0, 1.0, by=0.1), col='grey')
abline(v=seq(0.0, 1.0, by=0.1), col='grey')
```



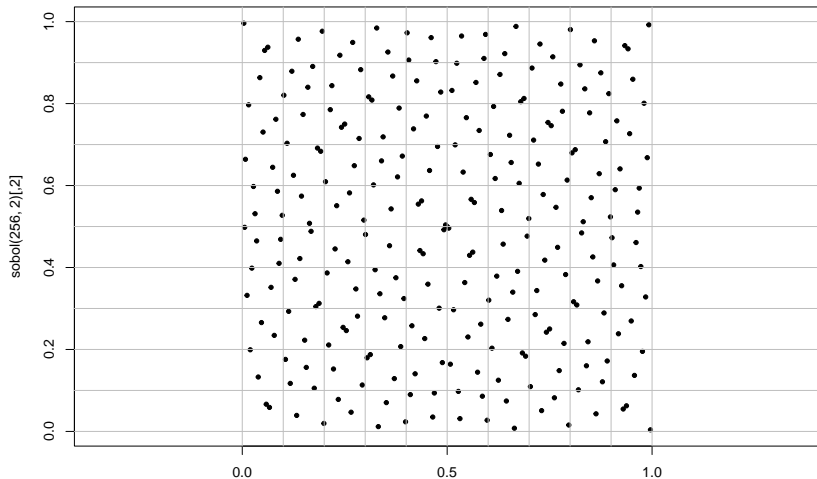
Distribution binomial $\text{Bi}(256, 0.01)$ des comptes

```
x = seq(0, 10)
y = dbinom(x, 256, 0.01)
plot(x, y)
```



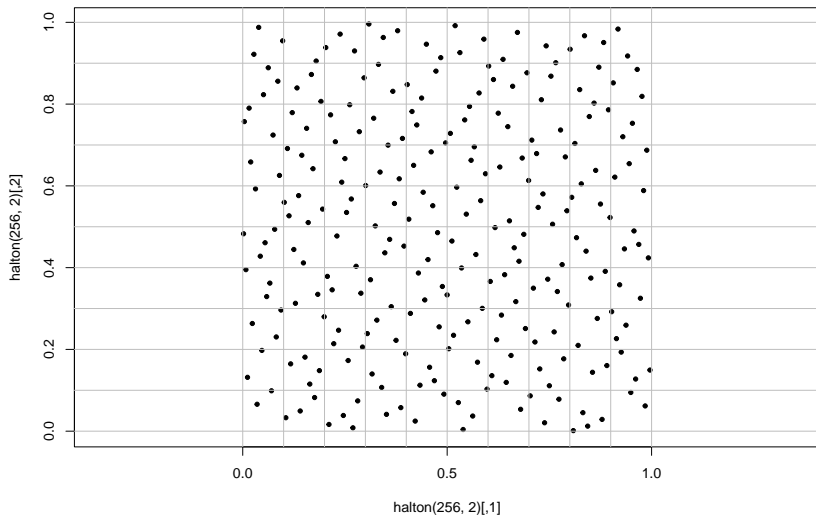
Suite de Sobol'

```
library(randtoolbox)
plot(sobol(256, 2), asp=1, pch=20, xlim=c(0,1), width=10, h
abline(h=seq(0.0, 1.0, by=0.1), col='grey')
abline(v=seq(0.0, 1.0, by=0.1), col='grey')
```



Suite de Halton

```
plot(halton(256, 2), asp=1, pch=20, xlim=c(0,1))  
abline(h=seq(0.0, 1.0, by=0.1), col='grey')  
abline(v=seq(0.0, 1.0, by=0.1), col='grey')
```



Méthodes pour les suites pseudo-aléatoires

- ▶ Il y a plusieurs **suites pseudo-aléatoires** proposées.
- ▶ Une méthode répandue est le **Mersenne Twister**.
- ▶ Les **Tests Diehard** est une suite de tests de plusieurs implications de l'hypothèse que $U_i \sim \text{iid } U(0, 1)$.
- ▶ Souvent, il y a un état x_i qui suit une règle déterministe $x_i = f(x_{i-1}, x_{i-2}, \dots)$ et une suite de nombres y_i (la suite sortante) qui suit une règle déterministe $y_i = g(x_i)$, où g est une fonction à sens unique qui rend infaisable le calcul de l'état à partir des observations des y_i .
- ▶ L'utilisateur peut initialiser l'état avec le choix d'une graine (seed), pour la reproductibilité.

```
set.seed(1234567890)
```

- ▶ Sinon, l'état est souvent une fonction de l'heure actuelle (en UNIX, le nombre de secondes depuis le début de 1/1/1970)

Variables (pseudo)-aléatoires non-uniformes

1. Méthode de l'inversion de la fonction de répartition
2. Méthode de rejet
3. Méthode de Ziggurat

Inversion de la fonction de répartition I

Théorème 2.1 de Devroye : Supposez que $F(x)$ est une fonction de répartition continue; son inverse F^{-1} est définie par

$$F^{-1}(u) = \inf \{x: F(x) = u\}, 0 \leq u \leq 1.$$

Alors

1. Si la variable aléatoire U suit la loi uniforme sur $[0, 1]$, la fonction de répartition de la variable aléatoire $X \equiv F^{-1}(U)$ est F .
2. Si la variable aléatoire X a F comme fonction de répartition, la loi de $F(X)$ est la loi uniforme sur $[0, 1]$.

La deuxième résultat est utilisé souvent en économétrie : transformation intégrale de probabilité, valeurs p .

Inversion de la fonction de répartition II

Preuve de 1 :

$$\begin{aligned}\Pr[F^{-1}(U) \leq x] &= \Pr[\inf\{y : F(y) = U\} \leq x] \\ &= \Pr[U \leq F(x)] = F(x).\end{aligned}$$

Preuve de 2 :

$$\Pr[F(X) \leq U] = \Pr[X \leq F^{-1}(u)] = F(F^{-1}(u)) = u.$$

Quelques exemples où l'inverse analytique est disponible

Loi	$F(x)$	$F^{-1}(u)$
Exponentiel	$1 - e^{-\lambda x}$	$\frac{1}{\lambda} \log(1 - u)$
Cauchy	$\frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{x - \bar{x}}{\sigma} \right)$	$\bar{x} + \tan \left(\pi \left(u - \frac{1}{2} \right) \right)$
Pareto, $x \geq b > 0$	$1 - \left(\frac{b}{x} \right)^a$	$\frac{b}{(1-u)^{1/a}}$
Weibull, $x \geq 0$	$1 - e^{-(x/\lambda)^k}$?

Lois non-uniformes par la méthode de rejet

- ▶ Deux densités de la méthode de rejet :
 - ▶ densité cible $f(x)$ (on veut simuler de cette loi)
 - ▶ densité de proposition $g(x)$ (tirer de cette loi est facile)
- ▶ Il faut que $g(x)$ domine $f(x)$ dans le sens que

$$\sup_x \frac{f(x)}{g(x)} \equiv M < \infty$$

- ▶ La méthode consiste en répétant les étapes suivantes jusqu'à une acceptation
 - ▶ tirer X de la loi avec densité $g(x)$
 - ▶ tirer U de la loi uniforme sur $[0, 1]$
 - ▶ accepter X comme un tirage de la loi cible si $U \leq f(X)/(g(X)M)$.

La méthode de rejet : probabilité d'acceptation

- ▶ La probabilité conditionnelle d'accepter, sachant X , est

$$\Pr \left[U \leq \frac{f(X)}{Mg(X)} \middle| X \right] = \frac{f(X)}{Mg(X)}.$$

- ▶ La probabilité inconditionnelle est (plus de rigueur [ici](#))

$$E_g \left[\frac{f(X)}{Mg(X)} \right] = \int \frac{f(x)}{Mg(x)} g(x) dx = \frac{1}{M} \int f(x) dx = \frac{1}{M}.$$

Exemple, loi gaussienne tronqué à $(-\infty, c]$, $c > 0$.

- Considérez la loi cible avec densité

$$f(x) = \frac{1}{\Phi(c)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} 1_{(-\infty, c]}(x),$$

où $c > 0$.

- C'est une loi gaussienne tronqué à la région $(-\infty, c]$.
- Notez que l'intégral de $f(x)$ est 1.
- Choisissez maintenant la loi $N(0, 1)$ comme loi de proposition :

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

- $M \equiv \sup f(x)/g(x) = 1/\Phi(c)$.
- Avec la méthode de rejet, on accepte X non-tronqué ssi $X < c$:

$$\frac{f(X)}{Mg(X)} = \begin{cases} 1 & X \leq c \\ 0 & X > c \end{cases}$$

Exemple, loi gaussienne tronqué à (c, ∞) , $c > 0$.

- ▶ On peut utiliser la $g(x)$, mais ici la probabilité d'acceptation inconditionnelle serait $1 - \Phi(c)$, qui peut être très petite.
- ▶ Première idée alternative : choisir une loi exponentielle avec taux λ et déplacement par c :

$$g(x) = \lambda e^{-\lambda(x-c)} 1_{(c, \infty)}(x).$$

- ▶ La dérivée de $\log f(x)$ est $-x$ et la dérivée de $\log g(x)$ est $-\lambda$.
- ▶ On met les deux en égalité à $x = c$ avec le choix $\lambda = c$.
- ▶ Le ratio $f(x)/g(x)$ est maximisé à $x = c^+$, où

$$M = \frac{f(c^+)}{g(c^+)} = \frac{e^{-c^2/2}}{\sqrt{2\pi}(1 - \Phi(c))c}.$$

- ▶ Pour $c = 2$, $M = 1.186608$ et la probabilité d'acceptation est $M^{-1} = 0.8427385$.

Première idée, code pour la graphique

```
# Point de troncation, probabilité de la région (c, infty)
c = 2
A = 1 - pnorm(c, 0, 1)

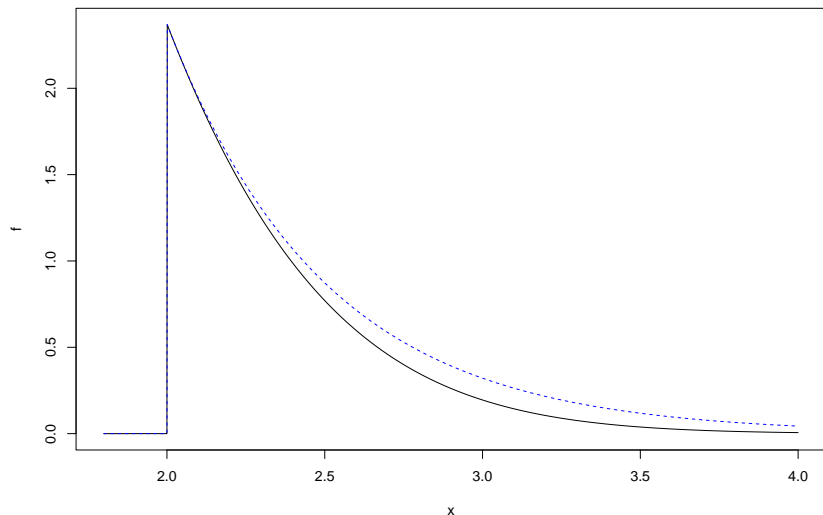
# Grille de points, densité cible (gaussienne tronquée)
x = seq(c-0.2, 4, by=0.001)
f = (dnorm(x) / A) * (x > c)

# Densité de proposition, M, aPr
lambda = c
g = lambda * exp(-lambda * (x-c)) * (x > c)
M = dnorm(c) / (A*lambda)
aPr = 1/M
c(M, aPr)

## [1] 1.1866078 0.8427385
```

Première idée, graphique

```
plot(x, f, type='l')  
lines(x, g*M, col='blue', lty='dashed')
```



Deuxième idée

- ▶ Retenons $g_\lambda(x) = \lambda e^{-\lambda(x-c)} 1_{(c,\infty)}$, mais cherchons la valeur optimal de λ .
- ▶ On peut écarter $\lambda < c$, où $f(x)/g_\lambda(x)$ est maximisée à $x = c$.
- ▶ Pour $\lambda > c$, $f(x)/g_\lambda(x)$ est maximisé au même point x^* que

$$\log f(x) - \log g_\lambda(x) = k - \frac{1}{2}x^2 + \lambda(x - c)$$

- ▶ La dérivée est $-x + \lambda = 0$ alors $f(x)/g_\lambda(x)$ est maximal à $x^* = \lambda$, peu importe la valeur de $\lambda > c$.
- ▶ Maintenant on calcule M comme fonction de λ :

$$\left. \frac{f(x)}{g_\lambda(x)} \right|_{x=\lambda} = \frac{e^{-\lambda^2/2}}{\sqrt{2\pi}(1 - \Phi(c))} \cdot \frac{1}{\lambda e^{-\lambda(\lambda-c)}} = \frac{e^{\lambda^2/2 - \lambda c}}{\sqrt{2\pi}(1 - \Phi(c))\lambda}$$

- ▶ Une condition de première ordre nécessaire pour un minimum :

$$\lambda - c - \frac{1}{\lambda} = 0.$$

- ▶ Une racine plus grand que c : $\lambda^* = (c + \sqrt{c^2 + 4})/2$

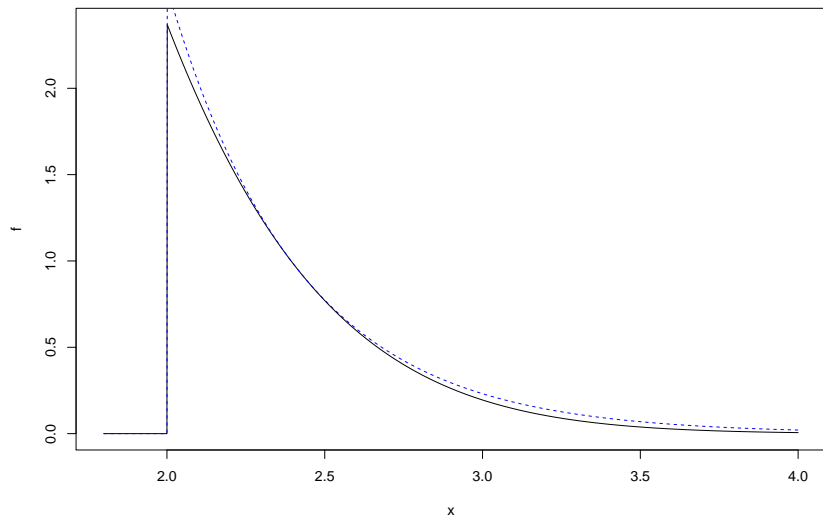
Deuxième idée, code

```
# Densité de proposition, M, aPr
lambda = (c + sqrt(4 + c^2))/2
g = lambda * exp(-lambda * (x-c)) * (x > c)
M = exp(lambda^2/2 - lambda*c) / (A*sqrt(2*pi)*lambda)
aPr = 1/M
c(M, aPr)

## [1] 1.0710705 0.9336453
```

Deuxième idée, graphique

```
plot(x, f, type='l')  
lines(x, g*M, col='blue', lty='dashed')
```



Deuxième idée, comparaison efficace de U et f/Mg

Rappel : pour $x \geq c$,

$$\frac{f(x)}{g(x)} = \frac{e^{-x^2/2}}{\sqrt{2\pi}(1 - \Phi(c))} \cdot \frac{1}{\lambda e^{-\lambda(x-c)}},$$

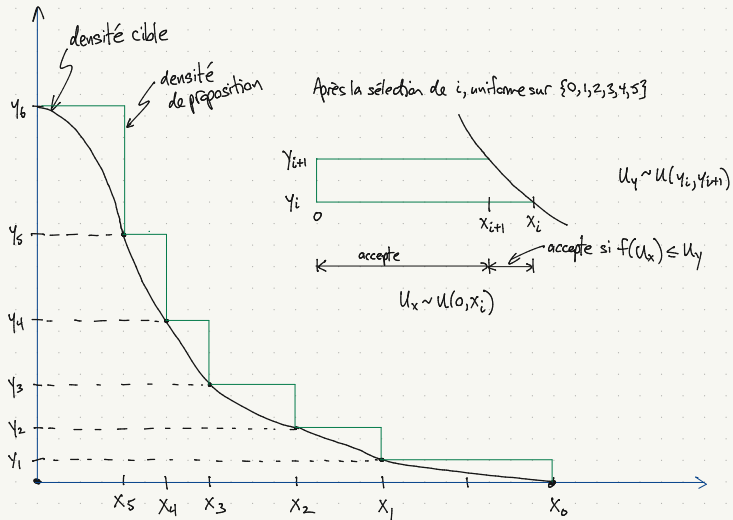
$$M = \frac{e^{\lambda^2/2 - \lambda c}}{\sqrt{2\pi}(1 - \Phi(c))\lambda}.$$

La probabilité conditionnelle d'acceptation $f(x)/(Mg(x))$ simplifie à

$$\frac{f(x)}{Mg(x)} = e^{-(x-\lambda)^2/2}.$$

Notez que $e^{-(x-\lambda)^2/2} \geq 1 - \frac{(x-\lambda)^2}{2}$, qui permet une acceptation rapide si $U \leq 1 - \frac{(x-\lambda)^2}{2}$, sans évaluer la fonction exponentielle.

La méthode Ziggurat



Commentaires sur la construction de la Ziggurat

- ▶ $n = 256$ niveaux est typique (8 bits aléatoire, car $256 = 2^8$).
- ▶ L'aire de la couche i , $(y_{i+1} - y_i)x_i = A$ est constante.
- ▶ L'aire constante donne $y_{i+1} = (A/x_i) + y_i$.
- ▶ L'inversion de $y_i = f(x_i)$ donne $x_i = f^{-1}(y_i)$ (la monotonie est importante).
- ▶ Si le support de la loi cible est $[0, \infty)$, la couche zéro est la région en dessous de $\min(y_1, f(x))$, qui doit avoir l'aire A .
- ▶ Le coût élevé de la construction de la Ziggurat est justifié s'il faut le faire seulement une fois pour tout.
- ▶ Convenable pour les lois $N(0, 1)$, $\text{Exp}(1)$, parce que
 - ▶ $N(\mu, \sigma)$ et $\text{Exp}(\lambda)$ sont des transformations simples de $N(0, 1)$ et $\text{Exp}(1)$.
 - ▶ La densité $\phi(x)$ de la loi $N(0, 1)$ est symétrique autour de zéro, montone pour $x > 0$
 - ▶ La densité de la loi $\text{Exp}(1)$ est monotone sur son support $[0, \infty)$.

Commentaires sur le tirage des variables aléatoires

- ▶ Supposez que la Ziggurat est déjà construite.
- ▶ La méthode de Ziggurat consiste en répétant les étapes suivantes jusqu'à une acceptation :
 - ▶ tirer l'index de la couche, de la loi uniforme sur $\{0, 1, \dots, n-1\}$.
 - ▶ tirer U_x de la loi uniforme sur $[0, x_i]$.
 - ▶ Si $U_x \leq x_{i+1}$, accepter U_x comme le tirage de la loi cible.
 - ▶ Sinon, tirer U_y de la loi uniforme sur $[y_i, y_{i+1}]$, accepter si $f(U_x) \leq U_y$.
- ▶ Si le support de la loi cible est $[0, \infty)$, le tirage de la couche zéro doit utiliser une autre méthode.
- ▶ L'aire constante des régions est importante pour un coût de tirage qui ne dépend pas de n .
- ▶ Pour n grand, on accepte avec haute probabilité sans tirer U_y ni évaluer $f(U_x)$.

Une Ziggurat (édifice religieux mésopotamien)

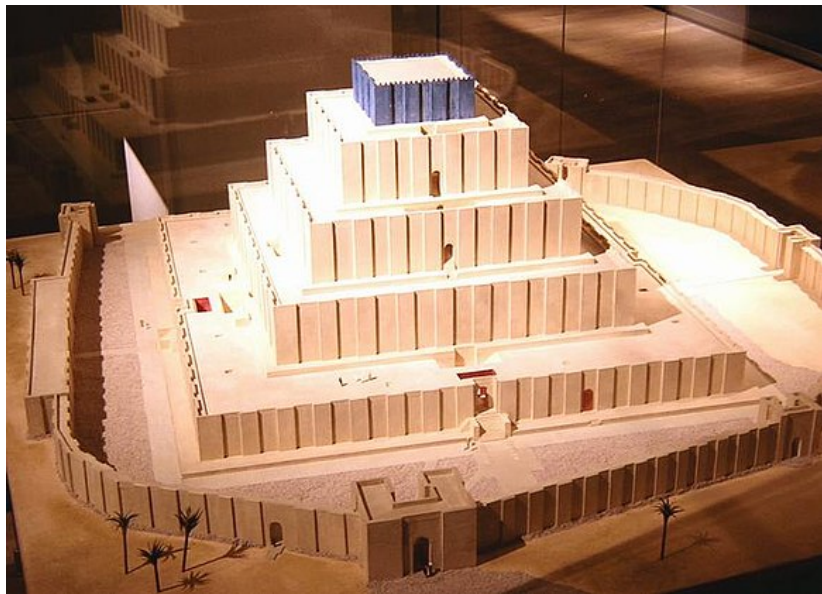


Figure 1: Ziggurat

Ziggurat de neige



Figure 2: Ziggurat de neige

Variables (pseudo)-aléatoires multivariée “faisables”

1. de la loi gaussienne multivariée (généralisation de la loi gaussienne)
2. de la loi Dirichlet (généralisation de la loi beta)
3. de la loi Wishart (généralisation de la loi gamma)

Tirer de la loi multivariée gaussienne

- ▶ Une transformation linéaire d'un vecteur de variables aléatoires gaussiennes
- ▶ Pour tirer $Y \sim N(\mu, \Sigma)$,
 - ▶ Tirer $X \sim N(0, I_n)$
 - ▶ Effectuer la décomposition de Cholesky $\Sigma = LL^\top$.
 - ▶ $Y = \mu + LX$.
- ▶ Variants :
 - ▶ loi uniforme sur la surface d'une hypersphère
 - ▶ loi uniforme sur une hypersphère
 - ▶ loi t de Student et autres mélanges de lois gaussiennes

Tirer de la loi Dirichlet

- ▶ Généralisation de la loi $\text{Be}(\alpha, \beta)$ (beta)
- ▶ Elle est une loi sur le n -simplexe Δ^n standard (ou de probabilité)

$$\Delta^n \equiv \left\{ x \in \mathbb{R}^n : x_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n x_i = 1 \right\}$$

- ▶ Pour tirer $Y \sim \text{Di}(\alpha_1, \dots, \alpha_n)$,
 - ▶ Tirer $X_i \sim \text{indep. Ga}(\alpha_i, \beta)$, $i = 1, \dots, n$.
 - ▶ Soit $X_{\text{tot}} = \sum_{i=1}^n X_i$.
 - ▶ $Y \equiv X_{\text{tot}}^{-1}(X_1, \dots, X_n) \sim \text{Di}(\alpha_1, \dots, \alpha_n)$.
- ▶ Variants :
 - ▶ Loi uniforme sur un simplexe standard ($\text{Di}(1, \dots, 1)$)
 - ▶ Loi uniforme sur un simplexe arbitraire par transformation linéaire

Tirer de la loi Wishart

- ▶ Généralisation de la loi $\text{Ga}(\alpha, \beta)$ (gamma)
- ▶ Elle est une loi sur l'espace de matrices symétriques et définies positives.
- ▶ Pour tirer $\Sigma \sim \text{Wi}(\nu, V)$, où V est une matrice définie positive,
 - ▶ Effectuer la décomposition de Cholesky $V = LL^\top$.
 - ▶ Tirer la matrice A : ses éléments sont indépendants et

$$A_{ij} = \begin{cases} N(0, 1) & i < j \\ \chi^2(\nu - i + 1) & i = j \\ 0 & i > j \end{cases}$$

- ▶ $\Sigma \equiv LAA^\top L^\top$ (LA est le facteur triangulaire inférieure de Σ)
- ▶ Variant : loi de Wishart inverse