SMEMP302

PROBABILISTIC COMPUTATIONAL METHODS
(2022 – 2023)

Chapter 1
Simulation of random variables

1.1. Pseudo-random numbers

Computers cannot generate a sequence of random number. We usually rely on « pseudo-random » numbers. It is considered as « good » sequence if it satisfies a serie of statistical tests (not described in this course).

Classic sequence:
$$\begin{cases} y_{n+1} \equiv a\, y_n + b \mod N \\ x_n = \frac{y_n}{N} \quad (y_n \in \{0, \dots, N-1\}) \end{cases}$$

where
$\gcd(a, N) = 1$ so that $\bar{a}$ (class of a modulo N) is invertible w.r.t. multiplication (modulo N)

Most simple case: $b \equiv 0$ → homogeneous congruent generator

<u>Remark:</u> We get a sequence (x_n) in $[0;1]$

Usually, we choose $N = p^\alpha$ with p a prime number.

1.2 The inverse distribution function method

Let μ be a probability distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with distribution function: $x \mapsto F(x) = \mu((-\infty, x])$.

Remember: F is right-continuous with left limit

F is nondecreasing

$\lim_{x \to +\infty} F(x) = 1$; $\lim_{x \to -\infty} F(x) = -1$

We define the canonical left inverse:

$$\forall u \in (0,1), \quad F_l^{-1}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}$$

We easily check that F_l^{-1} is non-decreasing, left-continuous and satisfies:

$$\forall u \in (0,1); \quad F_l^{-1}(u) \leq x \iff u \leq F(x)$$

Proposition: If $U \sim \mathcal{U}([0,1])$ then $X := F_l^{-1}(U) \sim \mu$.

Proof: We have $\{X \leq x\} = \{F_l^{-1}(U) \leq x\}$

$= \{U \leq F(x)\}$

so $\mathbb{P}(X \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$ $\square$

Remarks: * If F is increasing and continuous on $\mathbb{R}$ then F has an inverse, denoted by F^{-1}. In this case $F^{-1} = F_l^{-1}$.

* If μ has a prob. density f such that $\{x : f(x) = 0\}$ has an empty interior then $\int_{-\infty}^{x} f(t)\mu(dt) = F(x)$ is continuous and increasing.

Let $E = \{x_1, \dots, x_n\}$ s.t. $x_1 < x_2 < \dots < x_n$

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}$ a r.v. such that

$\forall k$, $\mathbb{P}(X = x_k) = p_k$ with all $p_k \in (0,1)$

and $p_1 + \dots + p_n = 1$

We get the distribution function: $F_X(x) = p_1 + \dots + p_i$

for $x \in [x_i, x_{i+1})$

F is nondecreasing

$\lim_{x \to +\infty} F(x) = 1$; $\lim_{x \to -\infty} F(x) = -1$

We define the canonical left inverse:

$\forall u \in (0,1)$, $F_\ell^{-1}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}$

<u>Examples</u> 1. Simulation of a Pareto (θ) $\theta > 0$ distribution.

We have (by definition) $\mathbb{P}_X(dx) = \frac{\theta}{x^{1+\theta}} \mathbb{1}_{x \geq 1}\, dx$

$\to$ distribution function $F_X(x) = (1 - x^{-\theta})\mathbb{1}_{x \geq 1}$

which is invertible : $F_X^{-1}(u) = (1-u)^{-1/\theta}$

so, if $U \sim \mathcal{U}([0;1])$, $(1-U)^{-1/\theta} \sim$ Pareto(θ)

as $U \overset{loi}{=} 1-U$, $U^{-1/\theta} \sim$ Pareto (θ)

2. Simulation of a distribution supported by a finite set

Let $E = \{x_1, \dots, x_n\}$ s.t. $x_1 < x_2 < \dots < x_n$

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}$ a r.v. such that

$\forall k$, $\mathbb{P}(X = x_k) = p_k$ with all $p_k \in (0,1)$

and $p_1 + \dots + p_n = 1$

We get the distribution function : $F_X(x) = p_1 + \dots + p_i$

for $x \in [x_i, x_{i+1})$

(We set $x_0 = -\infty$, $x_{n+1} = +\infty$.)

We get : $F_{X,l}^{-1}(u) = \sum_{k=1}^{m} x_k \mathbb{1}_{p_1+\dots+p_{k-1} < u \leq p_1+\dots+p_k}$

So $X \overset{\text{loi}}{=} \sum_{k=1}^{m} x_k \mathbb{1}_{p_1+\dots+p_{k-1} < U \leq p_1+\dots+p_k}$ with $U \sim \mathcal{U}([0,1])$

1.3 The acceptance/rejection method

Let μ be a non-negative σ-finite measure on $(E, \mathcal{E})$ (a measurable space). Let $f, g : (E, \mathcal{E}) \longrightarrow \mathbb{R}_+$ be two Borel functions (i.e. $f^{-1}(B) \in \mathcal{E}$ for all $B \in \mathcal{B}(\mathbb{R})$).
Assume that $f \in L^1(\mu)$ with $\int_E f \, d\mu > 0$ and g is a probability density w.r.t. μ, with $g > 0$ μ-a.s..
And assume $\exists\, c > 0$ such that:

$$f(x) \leq c\, g(x) \quad \mu(dx)\text{-a.s.}$$

Requirements for implementation :

* Numerical value of c is known.
* We know how to simulate a sequence of iid (Y_k) of law $g.\mu$.
* We can compute the ratio $\frac{f(x)}{g(x)}$ for all x.

Proposition : Let (U_n, Y_n) be a sequence of iid r.v. with law $\mathcal{U}([0,1]) \otimes \mathbb{P}_Y$ (independent marginals) defined on $(\Omega, \mathcal{A}, \mathbb{P})$ where $\mathbb{P}_Y(dy) = g(y)\mu(dy)$. Set

$$\tau := \min\{k \geq 1,\ c\, U_k\, g(Y_k) \leq f(Y_k)\}$$

then $\tau \sim \mathcal{G}(p)$ with $p = \mathbb{P}(c\, U_1\, g(Y_1) \leq f(Y_1)) = \frac{\int_E f \, d\mu}{c}$ and $X := Y_\tau$ has law ν

$\downarrow$

$$\nu = \frac{f}{\int_E f \, d\mu} . \mu$$

<u>Proof:</u> Let $(U,Y) \sim \mathcal{U}([0;1]) \otimes \mathbb{P}_Y$.

For all h bounded Borel test function, we compute:

$$\mathbb{E}\left(h(Y)\,\mathbb{1}_{cUg(Y)\le f(Y)}\right) = \int_{E\times[0,1]} h(y)\,\mathbb{1}_{cu\,g(y)\le f(y)}\; g(y)\,\mu(dy)$$

Fubini $\rightarrow$ $$= \int_E \left[\int_0^1 \mathbb{1}_{cu\,g(y)\le f(y)}\,du\right] h(y) \times g(y)\,\mu(dy)$$

as $\frac{f(y)}{c\,g(y)} \le 1$ $\rightarrow$ $$= \int_E \frac{f(y)}{c\,g(y)} \times h(y)\,g(y)\,\mu(dy)$$

$$= \frac{1}{c}\int_E f(y)\,h(y)\,\mu(dy)$$

With $h=1$, we get $p = \mathbb{P}(c\,Ug(Y)\le f(Y)) = \frac{1}{c}\int_E f(y)\,\mu(dy)$

$$\overline{\int_E f(y)\,\mu(dy)}$$

$$= \int_E h(y)\,\nu(dy)$$ $\square$

<u>Application:</u> uniform distribution on a bounded domain $D \subset \mathbb{R}^d$.

Let $D \subset a + [-M;M]^d$, $\lambda_d(D) > 0$ (λ_d is the Lebesgue measure, $a \in \mathbb{R}^d$, $M > 0$)

and let $Y = \mathcal{U}(a + [-M,M]^d)$,

let $\tau_D := \min\{n : Y_n \in D\}$ where (Y_n) is i.i.d. with same distrib. as Y. Then $Y_{\tau_D} \sim \mathcal{U}(D)$ $\rightarrow$ uniform distribution on D

<u>Proof:</u> Let $(U, Y) \sim \mathcal{U}([0;1]) \otimes \mathbb{P}_Y$.

For all h bounded Borel test function, we compute:

$$\mathbb{E}(h(Y_\tau)) = \sum_{n\geq 1} \mathbb{E}(\mathbb{1}_{\tau=n} \, h(Y_n))$$

$$= \sum_{n\geq 1} \mathbb{P}(c\,U_1 g(Y_1) > f(Y_1))^{n-1} \, \mathbb{E}(h(Y_1)\, \mathbb{1}_{c\,g(Y_1) U_1 < f(Y_1)})$$

$$= \sum_{n\geq 1} (1-p)^{n-1} \, \mathbb{E}(h(Y_1)\, \mathbb{1}_{c U_1 g(Y_1) < f(Y_1)})$$

$$= \frac{1}{p} \int_E \frac{f(y)\, h(y)}{c} \mu(dy)$$

$$= \frac{\int_E f(y)\, h(y)\, \mu(dy)}{\int_E f(y)\, \mu(dy)}$$

$$= \int_E h(y)\, \nu(dy) \qquad \square$$

<u>Application:</u> uniform distribution on a bounded domain $D \subset \mathbb{R}^d$.

Let $D \subset a + [-M, M]^d$, $\lambda_d(D) > 0$ (λ_d is the Lebesgue measure, $a \in \mathbb{R}^d$, $M > 0$)
and let $Y = \mathcal{U}(a + [-M, M]^d)$,
let $\tau_D := \min\{n : Y_n \in D\}$ where (Y_n) is i.i.d. with same distrib. as Y. Then $Y_{\tau_D} \sim \mathcal{U}(D)$ → uniform distribution on D

This follows from the above Proposition with

$E = a + [-M, M]^d$, $\mu := \lambda_d / _{a+[-M,M]^d}$

(Lebesgue measure on $a + [-M, M]^d$)

and $\begin{cases} g(u) = (2M)^{-d} \mathbb{1}_{a+[-M,M]^d}(u) \\ f(x) = \mathbb{1}_D(x) \leq \underbrace{(2M)^d}_{=:c} g(x) \end{cases}$

so that $\dfrac{f}{\int_{\mathbb{R}^d} f \, d\mu} \cdot \mu = \mathcal{U}(D)$

As a matter of fact, with the notation of the above proposition : $\tau = \min\{k \geq 1, c U_k \leq \frac{f}{g}(Y_k)\}$

As $\frac{f}{g}(y) > 0 \iff y \in D$ (in which case $\frac{f}{g}(y) = 1$)

So $\tau = \tau_D$.

1.4 Simulation of Gaussian random vectors

1.4.1. d- dim. standard normal vectors

Prop. (Box-Müller method)

Let $R^2 \sim \mathcal{E}(\frac{1}{2})$ and Θ (ind. of R^2) $\sim \mathcal{U}([0, 2\pi])$. Then

$X := (R\cos\Theta, R\sin\Theta) \overset{law}{=} \mathcal{N}(0, I_2)$

(where $R = \sqrt{R^2}$)

Proof : Let f be a bounded Borel function

$$\iint_{\mathbb{R}^2} f(x_1, x_2) \exp\left(-\frac{(x_1^2 + x_2^2)}{2}\right) \frac{dx_1 \, dx_2}{2\pi} = (*)$$

$$\parallel \quad \begin{cases} x_1 = \rho\cos\theta \\ x_2 = \rho\sin\theta \end{cases}$$

$$\iint_{\mathbb{R}^2} f(\rho\cos\theta, \rho\sin\theta) e^{-\rho^2/2} \mathbb{1}_{\mathbb{R}_+^*}(\rho) \mathbb{1}_{[0,2\pi]}(\theta) \rho \frac{d\rho \, d\theta}{2\pi}$$

because $\begin{bmatrix} \frac{\partial x_1}{\partial \rho} & \frac{\partial x_2}{\partial \rho} \\ \frac{\partial x_1}{\partial \theta} & \frac{\partial x_2}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\rho\sin\theta & \rho\cos\theta \end{bmatrix}$

→ determinant $\rho\cos^2\theta + \rho\sin^2\theta = \rho$

while $(x_1,x_2) \in \mathbb{R}^2\setminus\{0\}$, $(\rho,\theta) \in \mathbb{R}_+^* \times [0,2\pi[$

and this is a bijection

so $dx_1 dx_2$ transforms into $|\rho|\, d\rho\, d\theta$
$$\overset{\shortparallel}{\rho\, d\rho\, d\theta}$$
(because $\rho \in \mathbb{R}_+$)

And so : $(*) = \mathbb{E}(f(\sqrt{R^2}\cos\theta, \sqrt{R^2}\sin\theta))$
$= f(X)$

Corollary (Box-Muller method) If $(U_1, U_2) \otimes \mathcal{U}([0,1])^{\otimes 2}$
$X := (\sqrt{-2\log U_1}\cos(2\pi U_2), \sqrt{-2\log U_1}\sin(2\pi U_2))$
Then $X \sim \mathcal{N}(0, I_2)$

Proof: Let f be a bounded Borel function:

$$\mathbb{E}(f(-2\log U_1)) = \int_0^1 f(-2\log u)\,du$$

$$\begin{cases} x = -2\log u \\ u = e^{-1/2 x} \end{cases} = \int_{+\infty}^{0} f(x)\, e^{-x/2}\left(-\frac{1}{2}\right) dx$$

$\hookrightarrow du = -\frac{1}{2} e^{-x/2} dx$

$$= \int_0^{+\infty} f(x)\, \frac{e^{-x/2}}{2}\, dx$$

so $-2\log U_1 \sim \mathcal{E}\left(\frac{1}{2}\right)$

1.4.2. Correlated d-dimensional Gaussian vectors, Gaussian processes

Let $X = (X^1, \dots, X^d)$ be a centered $\mathbb{R}^d$-valued Gaussian vector with covariance matrix $\Sigma_1 = (\operatorname{Cov}(X_i, X_j))_{1\leq i,j\leq d}$

The symmetric matrix Σ_1 is positive semidefinite (meaning : $\forall$ vector V, $V^T \Sigma_1 V \geq 0$).

« Gaussian vector » means that : $\forall u \in \mathbb{R}^q$, $\sum_{i=1}^{d} u_i X^i$ is Gaussian with variance $u^T \Sigma_1 u$

Lemma : Let Y be a $\mathbb{R}^d$-valued square integrable random vector and let $A \in \mathcal{M}(q, d)$ ($q \times d$ matrix). Then the covariance matrix C_{AY} of the random vector AY is given by $C_{AY} = A \underbrace{C_Y}_{\text{covariance of } Y} A^T$

Proof: $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_d \end{bmatrix}$ $\quad A = ((a_{ij}))_{\substack{1 \leq i \leq q \\ 1 \leq j \leq d}}$

$$\left(\sum_{j=1}^{d} a_{ij} (Y_j - \mathbb{E}(Y_j)) \right)^2 = \sum_{j=1}^{d} \sum_{k=1}^{d} a_{ij} a_{ik} (Y_j - \mathbb{E}(Y_j))(Y_k - \mathbb{E}(Y_k))$$

$\hookrightarrow \mathbb{E}(\ldots) = \sum_{j=1}^{d} \sum_{k=1}^{d} a_{ij} a_{ik} (C_Y)_{jk} = (A C_Y A^T)_{ij}$ $\square$

Application : When the covariance matrix Σ_1 is invertible, we have the Choletski decomposition :

$$\Sigma_1 = T T^T \quad \text{(where } T \text{ is a lower triangular matrix, i.e. } T_{ij} = 0 \text{ if } i < j)$$

(This is a consequence of the Hilbert - Schmidt orthonormalisation procedure)

and there exists a unique such lower triangular matrix with diagonal terms $T_{ii} > 0$ ($1 \leq i \leq d$). Owing to the above Lemma, $TZ \sim \mathcal{N}(0 ; \Sigma_1)$

Application to the simulation of a standard Brownian motion at fixed times

Let $(W_t)_{t\geq 0}$ be a standard Brownian motion defined on $(\Omega, \mathcal{B}, \mathbb{P})$. By def. of a Brownian motion:

$$\begin{cases} \mathbb{E}(W_s W_t) = s \wedge t & (\forall s, t) \\ \mathbb{E}(W_t) = 0 & (\forall t) \end{cases}$$

and (for $t_1 < t_2 < \dots < t_{n-1} < t_n$)

$$(W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}) \sim \mathcal{N}(0, \mathrm{Diag}(t_1, t_2 - t_1, \dots, t_n - t_{n-1}))$$

(because the increments of W are ind. and stationary)

So
$$\begin{bmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ \vdots \\ W_{t_n} - W_{t_{n-1}} \end{bmatrix} \overset{law}{=} \mathrm{Diag}(\sqrt{t_1}, \sqrt{t_2 - t_1}, \dots, \sqrt{t_n - t_{n-1}}) \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

where $(Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$

$\downarrow$

Proof in detail:

$$\begin{bmatrix} W_{t_1} \\ \vdots \\ W_{t_n} \end{bmatrix} = L \begin{bmatrix} W_{t_1} \\ \vdots \\ W_{t_n} - W_{t_{n-1}} \end{bmatrix} \quad \text{where } L = ((\mathbb{1}_{i \geq j}))_{1 \leq i,j \leq n}$$

We set $T = L\,\mathrm{Diag}(\sqrt{t_1}, \sqrt{t_2 - t_1}, \dots, \sqrt{t_n - t_{n-1}})$ and we can check that $T = ((\sqrt{t_j - t_{j-1}}\, \mathbb{1}_{i \geq j}))_{1 \leq i,j \leq n}$ and that

$$\begin{bmatrix} W_{t_1} \\ \vdots \\ W_{t_n} \end{bmatrix} = T \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}.$$

So T is the Cholevski decomposition of $\Sigma^W_{t_1, \dots, t_n}$