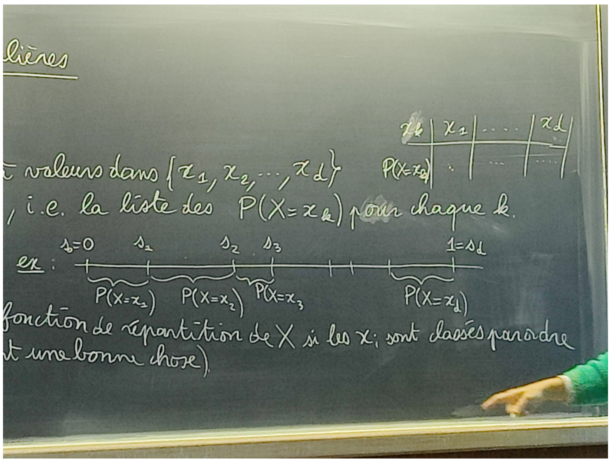
Simulation : Chapitre 3 : Simulation de lois particulières

Charles Vin

2021

1 Simulations de lois discrète

On veut simuler une v.a. X à valeur dans $\{x_1, x_2, ..., x_d\}$.
On connait son tableau de loi, i.e. la liste des $P(X = x_k)$ pour chaque k
On note $s_k = \sum_{i=1}^{k} P(X = x_i)$. Les s_k sont les valeurs prises par la fonction de répartition de X si les x_i sont classés par ordre croissant (ce qui n'est pas forcément une bonne chose). (exemple Figure 1)



On tire $U \sim unif(]0;1[)$: on obtient $U(w)$. On détermine l'unique k tel que $s_{k-1} < U(w) \leq s_k$ et on pose $Y(w) = x_k$.
On a alors

$$\begin{aligned} P(Y = x_k) &= P(s_{k-1} < U \leq s_k) \\ &= s_k - s_{k-1} = P(X = x_k) \end{aligned}$$

Donc Y a même loi que X

Efficacité :
On test successivement si $U \leq s_1$, si $U \leq s_2$ donc l'algorithme se termine plus vite si on classe les x_k par ordre de probabilités décroissantes :

$$P(X = x_1) \geq P(X = x_2) \geq \cdots \geq P(X = x_d).$$

Exemple 1.1. On lance un palet sur la Table 1.1. X indique le nombre de points obtenus par un lancer au hasard. $X(\Omega) = \{0, 5, 10\}$. La loi de X dans la Table 1.1.

5	5	5
5	0	5
0	10	0
5	0	5
5	5	5

On tire $U \sim Unif(]0;1[)$
Si $U < \frac{10}{15}$ alors $Y = 5$ sinon si $U < \frac{14}{15}$ alors $Y = 0$ sinon $Y = 10$ (voir Figure 1)

x_k	0	5	10
$P(X = x\_k)$	4/15	10/15	1/15

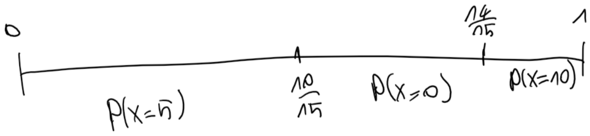


Figure 1 – Exemple

2 Technique spécifiques pour des lois discrètes particulière

2.1 Loi de Bernouilli

Si $U \sim Unif(]0;1[)$, $\mathbb{1}_{U \leq p} \sim Ber(p)$ (voir Figure 2)

Figure 2 – Technique par fonction quantiles j'ai pas l'image sorry

2.2 Loi binomiales et multinomiales

Si $U_1, U_2, ..., U_n$ sont i.i.d. de loi $Unif(]0;1[)$ alors $\sum_{i=1}^{n} \mathbb{1}_{U_i \leq p} \sim Bin(n,p)$

Note. **Rappel :** La loi multinomiales de paramètres $(n, p_1, p_2, ..., p_k)$ où les p_i sont positifs et de somme 1 est la loi du vecteur aléatoire $(X_1, X_2, ..., X_k)$ représentant le nombre de résultats de type obtenus en n expérience indépendantes ayant toutes une probabilité p_1 de donner le résultat de type 1, une probabilité p_2 de donner le résultat de type 2, etc.

$(X, n-X) \sim Multinomiales(n, p, 1-p)$ si $X \sim Bim(n,p)$

La même technique permet de simuler les lois multinomiales

Si $U_1, \ldots, U_n$ i.i.d. $Unif(]0;1[)$

$$(\sum_{i=1}^{n} \mathbb{1}_{U \leq p_1}, \sum_{i=1}^{n} \mathbb{1}_{p_1 < U \leq p_1 + p_2}, \sum_{i=1}^{n} \mathbb{1}_{p_1 + p_2 < U \leq p_1 + p_2 + p_3}, \ldots, \sum_{i=1}^{n} \mathbb{1}_{p_1 + \cdots + p_{k-1} < U \leq 1}) \sim Mult(n, p_1, \ldots, p_k).$$

2.3 Loi géométrique

Rappel : La loi $Gom(p)$ est la loi du nombre de tentatives indépendantes nécessaire pour obtenir le premier succès si chaque tentative a une proba p de succès

Si on simule la loi géométrique de paramètre p par cette méthode, le nombre moyen de v.a. Uniforme sur O et 1 nécessaire pour obtenir un tirage de $X \sim Geom(p)$ est $E(X) = \frac{1}{p}$ qui est très grand si p est très petit. Il est préférable surtout pour p petit, d'utiliser une méthode qui ne nécessite pas d'un tirage de loi uniforme.

Proposition 2.1. *Si $V \sim Exp(a)$ alors $\lceil V \rceil \sim Geom(1 - e^{-a})$ où $\lceil \rceil$ est la partie entière supérieure :* $\lceil x \rceil = \min\{k \in \mathbb{Z}, x \leq k\}$

Preuve :

$$\begin{aligned} P(\lceil V \rceil = k) = P(k-1 < V \leq k) &= \int_{k-1}^{k} a e^{-ax} dx \\ &= [-e^{-ax}]_{k-1}^{k} \\ &= -e^{-ak} + e^{-a(k-1)} \\ &= e^{-a(k-1)}(1 - e^{-a}) \\ &= p(1-p)^{k-1} \end{aligned}$$

donc $\lceil V \rceil \sim Geom(1-e^{-a})$. Suite :

$$1 - e^{-a} = p \Leftrightarrow e^{-a} = 1 - p \Leftrightarrow a = -\ln(1-p).$$

On sait simuler $Exp(-\ln(1-p))$ par $\frac{-\ln U}{-\ln(1-p)} = \frac{\ln U}{\ln(1-p)}$ où $U \sim Unif(]0;1[)$
CCL : Simulation de Géom(p) :

$$\lceil \frac{\ln U}{\ln(1-p)} \rceil \sim Geom(p) \text{ si } U \sim Unif(]0;1[).$$

$\square$

2.4 Loi de Poisson

Proposition 2.2. *Le nombre maximal de v.a. exponentielles indépendantes de paramètre a qu'on peut additionner sans dépasser 1 suit la loi de Poisson de paramètre a.*
Si $X_1, X_2, \dots$ sont i.i.d. $Exp(a)$

$$\forall n \in \mathbb{N}, P(\sum_{i=1}^{n} X_i \leq 1 < \sum_{i=1}^{n+1} X_i) = e^{-a}\frac{a^n}{n!}.$$

Conséquence : Si $U_1, U_2, \dots$ i.i.d $Unif(]0;1[)$ alors l'entier aléatoire N tel que $\prod_{i=1}^{N} U_i \geq e^{-a} > \prod_{i=1}^{N+1} U_i$ suit la loi $Pois(a)$

Pourquoi ?

$$P(\sum_{i=1}^{n} X_i \leq 1 < \sum_{i=1}^{n+1} X_i)$$
$$= \int_{\mathbb{R}^{n+1}} \mathbb{1}_{\sum_{i=1}^{n} X_i \leq 1 < \sum_{i=1}^{n+1} X_i)} \prod_{i=1}^{n+1} (ae^{-ax_i}\mathbb{1}_{x_i>0}) dx_1 \dots dx_{n+1}$$
$$= a^{n+1} \int_{\mathbb{R}^{n+1}} \mathbb{1}_{x_1+\dots+x_n \leq 1, x_1+\dots+x_n+x_{n+1}>1, x_1>0 \dots x_{n+1}>0} * e^{-a(x_1+\dots+x_{n+1})} dx_1 \dots dx_{n+1}$$

Changement de variable :

$$\begin{aligned} u_1 = x_1 &\Leftrightarrow x_1 = u_1 \\ u_2 = x_1 + x_2 &\Leftrightarrow x_2 = u_2 - u_1 \\ u_3 = x_1 + x_2 + x_3 &\Leftrightarrow x_3 = u_3 - u_2 \\ &\dots \\ 1 + \dots + x_n &\Leftrightarrow x_n = u_n - u_{n-1} \\ u_{n+1} = x_1 + \dots x_{n+1} &\Leftrightarrow x_{n+1} = u_{n+1} - u_n \end{aligned}$$

$$dx_1 \dots dx_{n+1} = \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \dots & \frac{\partial x_1}{\partial u_{n+1}} \\ \frac{\partial x_2}{\partial u_1} & \dots & \frac{\partial x_2}{\partial u_{n+1}} \\ \vdots & \dots & \vdots \\ \frac{\partial x_{n+1}}{\partial u_1} & \dots & \frac{\partial x_{n+1}}{\partial u_{n+1}} \end{pmatrix} \right| = \begin{pmatrix} 1 & 0 & \dots & 0 & \\ -1 & 1 & \ddots & \vdots & \\ 0 & \ddots & \ddots & 0 & \\ \vdots & \ddots & \ddots & 1 & \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}.$$

<u>Nouveau cours du 23/11</u>
C'est le chaos cette matrice, j'arrive pas à la faire en latex

$$= |\det Jac(u)| \, du_1 \dots du_{n+1} = |1| \, du_1 \dots du_{n+1}.$$

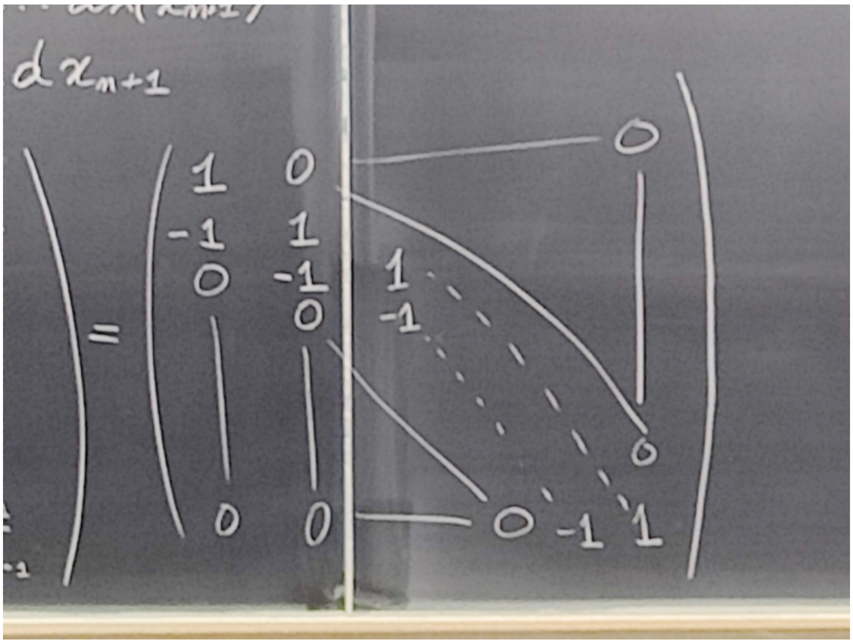


Figure 3 – <caption>

$$
\begin{aligned}
P(\sum_{i=1}^{n} X_i \leq 1 < \sum_{i=1}^{n+1} X_i) &= a^{n+1} \int_{\mathbb{R}^{n+1}} \mathbb{1}_{u_n \leq 1} \mathbb{1}_{u_{n+1}>1} e^{-a u_{n+1}} \mathbb{1}_{u_{n+1}>u_n>\cdots>u_1>0} \, |1| \, du_1 \ldots du_{n+1} \\
&= a^{n+1} \int_{\mathbb{R}^n} (\int_{\mathbb{R}} \mathbb{1}_{u_n<1} \mathbb{1}_{u_{n+1}>1} e^{-a u_{n+1}} \mathbb{1}_{u_{n+1}>u_n} \mathbb{1}_{u_n>u_{n-1}>\cdots>u_1>0} () du_{n+1}) du_1 \ldots du_n \\
&= a^{n+1} \int_{\mathbb{R}^n} \mathbb{1}_{u_n<1} \mathbb{1}_{u_n>u_{n-1}>\cdots>u_1>0} (\int_{\mathbb{R}} \mathbb{1}_{u_{n+1}>1} e^{-a u_{n+1}} du_{n+1}) du_1 \ldots du_n \\
&= \frac{a^{n+1}}{a} e^{-a} \int_{\mathbb{R}^n} \mathbb{1}_{1 \geq u_n \geq u_{n-1}>\cdots>u_1>0} du_1 \ldots du_n \\
&= a^n e^{-a} \int_{\mathbb{R}^n} \mathbb{1}_{u_1<u_2<\cdots<u_n} (\prod_{i=1}^{n} \mathbb{1}_{0<u_i<1}) du_1 \ldots du_n \\
&= a^n e^{-a} P(U_1 < U_2 < \cdots < U_n) \\
&= a^n e^{-a} P(U_1 < \cdots < U_n) \text{ pour } U_1, \ldots, U_n \text{ i.i.d. } Unif(]0;1]) \qquad = a^n e^{-a} \frac{1}{n}
\end{aligned}
$$

CCL : Si les X_i sont i.i.d. $Exp(a)$ alors le nombre aléatoire N de X_i tel que $\sum_{i=1}^{N} X_i \leq 1 < \sum_{i=1}^{N+1} X_i$ satifait $P(N=n) = \frac{a^n}{n!} e^{-a} \Leftrightarrow N \sim Pois(a)$

On sait que si $U_1, \ldots, U_n$ i.i.d. $Unif(]0;1[)$ alors

$$X_1 \frac{-\ln U_1}{a}, X_2 \frac{-\ln U_2}{a} \text{ i.i.d. } Exp(a).$$

$N \sim Pois(a)$ si N est déterminé par

$$
\begin{aligned}
&\sum_{i=1}^{N} \frac{-\ln U_i}{a} \leq 1 < \sum_{i=1}^{N+1} \frac{-\ln U_i}{a} \\
&-\frac{1}{a} \ln(\prod_{i=1}^{N} U_i) \leq 1 < -\frac{1}{a} \ln(\prod_{i=1}^{N+1} U_i) \\
&\prod_{i=1}^{N} U_i \geq e^{-a} > \prod_{i=1}^{N+1} U_i
\end{aligned}
$$

2.5 Simulation de gaussiennes

Rappel : Si X simule la $\mathcal{N}(0,1)$ alors $\sigma X + m$ simule $\mathcal{N}(m, \sigma^2)$
Inversion de la fonction de répartition : La fonction quantile de la $\mathcal{N}(0,1)$ est

$$\alpha \mapsto \inf\{t \in \mathbb{R}, \int_{\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \geq \alpha\}.$$

La méthode de simulation par inversion de la fonction de répartition n'est pas pratique ici
Méthode du rejet : Algorithme de Ziggurat, on l'utilisait sur les anciens processeurs, pas pratique
Méthode de Box-Muller : Si U et V sont indépendantes de loi $Unif(]0;1[)$ alors

$$\begin{aligned} X &= \sqrt{-2\ln U}\cos(2\pi V) \\ Y &= \sqrt{-2\ln U}\sin(2\pi V) \end{aligned}$$

sont indépendantes de loi $\mathcal{N}(0,1)$

Preuve : Calcul de la fonction de répartion de X, Y.
Vérifions que $\forall s, t \in \mathbb{R}$

$$P(X \leq s, Y \leq t) = \int_{\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx * \int_{\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

$$\begin{aligned} P(X \leq s, Y \leq t) &= P(\sqrt{-2\ln U}\cos(2\pi V) \leq s \text{ et} \sqrt{-2\ln U}\sin(2\pi V) \leq t) \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\sqrt{-2\ln U}\cos(2\pi V) \leq s} \mathbb{1}_{\sqrt{-2\ln U}\sin(2\pi V) \leq t} \mathbb{1}_{0<u<1} \mathbb{1}_{0<v<1} dudv \end{aligned}$$

Note. $\mathbb{1}_{0<u<1}\mathbb{1}_{0<v<1}$ est le produit des densités de U et V.
Changement de variable :

$$\begin{aligned} &x = \sqrt{-2\ln u}\cos(2\pi v), u = e^{-\frac{x^2+y^2}{2}} \\ &y = \sqrt{-2\ln u}\sin(2\pi v), v = \\ &x^2 + y^2 = -2\ln u \Leftrightarrow -\frac{x^2+y^2}{2} = \ln u \\ &\frac{y}{x} = \tan(2\pi v) \end{aligned}$$

$$]0;1[^2 \to \mathbb{R}^2 \setminus (]0;+\infty[\times\{0\}) \qquad (u,v) \mapsto (x,y)$$

La jacobienne :

$$\begin{aligned} dxdy &= \left|\det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}\right| dudv \\ Jac(u,v) &= \begin{pmatrix} \cos(2\pi v)(\frac{-2*\frac{1}{u}}{2\sqrt{-2\ln u}}) & -2\pi\sin(2\pi v)\sqrt{-2\ln u} \\ \sin(2\pi v)(\frac{-2*\frac{1}{u}}{2\sqrt{-2\ln u}}) & 2\pi\cos(2\pi v)\sqrt{-2\ln u} \end{pmatrix} \\ \det Jac(u,v) &= \frac{-1}{u\sqrt{-2\ln u}} * 2\pi\sqrt{-2\ln u}\det\begin{pmatrix} \cos(2\pi v) & -\sin(2\pi v) \\ \sin(2\pi v) & \cos(2\pi v) \end{pmatrix} \\ &= \frac{-2\pi}{u} * 1 \\ dxdy &= \frac{2\pi}{u} dudv \\ dudv &= \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \end{aligned}$$

$$
\begin{aligned}
P(X \leq s, Y \leq t) &= \int_{\mathbb{R}^2 \setminus (]0;+\infty[\times\{0\})} \mathbb{1}_{x \leq s} \mathbb{1}_{y \leq t} \frac{1}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} dxdy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{x \leq s} \mathbb{1}_{y \leq t} \frac{1}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} dxdy \\
&= \int_{\mathbb{R}} \mathbb{1}_{y \leq t} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} (\int_{\mathbb{R}} \mathbb{1}_{x \leq s} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx) dy \\
&= \phi(s)\phi(t)
\end{aligned}
$$

$\square$