Modélisation de suites de variables aléatoires indépendantes

Références : Ouvrard, *Probabilités 2*, p 53-59

Soit $x \in [0,1[$, on pose les suites $(D_n(x))_{n\geqslant 1}$ et $(R_n(x))_{n\geqslant 0}$ définies par

$$\left\{\begin{array}{l} R_0(x) = x \\ Dn(x) = \lfloor 2R_{n-1}(x) \rfloor \\ R_n(x) = 2R_{n-1}(x) - D_n(x) \end{array}\right. .$$

Elles vérifient $D_n \in \{0,1\}$ et $R_n \in [0,1[$.
De plus, par récurrence, on a pour tout $n \in \mathbb{N}$

$$x = \sum_{j=1}^{n} \frac{D_j(x)}{2^j} + \frac{1}{2^n} R_n(x) = \sum_{j=1}^{\infty} \frac{D_j(x)}{2^j}.$$

On a ainsi défini le développement dyadique de $x \in [0,1[$.
Attention, celui-ci est non-unique. Par exemple, $\frac{1}{2} = \sum_{j=2}^{\infty} \frac{1}{2^j}$.

Théorème.

On se place sur $([0,1[, \mathcal{B}([0,1[), \mathbb{P})$ avec $\mathbb{P}$ la restriction de la mesure de Lebesgue sur $[0,1[$.
La suite $(D_n)_n$ est une suite de variables aléatoires iid de loi $b(1/2)$. De plus, pour tout $n \in \mathbb{N}$, R_n suit une loi uniforme sur $[0,1[$ et les variables aléatoires D_1, ..., D_n et R_n sont mutuellement indépendantes.

Démonstration. • Soit $n \in \mathbb{N}^*$ et $\varepsilon^n = (\varepsilon_1, ..., \varepsilon_n) \in \{0,1\}^n$, on note

$$I_{\varepsilon^n}^n = \left[\sum_{j=1}^{n} \frac{\varepsilon_j}{2^j}, \sum_{j=1}^{n} \frac{\varepsilon_j}{2^j} + \frac{1}{2^n} \right[.$$

On remarque que les $I_{\varepsilon^n}^n$ forment une partition de $[0,1[$ et que les D_j sont constants dessus.
On a

$$\bigcap_{j=1}^{n} \{D_j = \varepsilon_j\} = I_{\varepsilon^n}^n,$$

donc

$$\mathbb{P}\left(\bigcap_{j=1}^{n} \{D_j = \varepsilon_j\} \right) = \frac{1}{2^n}.$$

Soit J une sous partie finie de $\{1, ..., n\}$, alors en sommant sur J^c, on a

$$\mathbb{P}\left(\bigcap_{j\in J} \{D_j = \varepsilon_j\} \right) = \frac{1}{2^{|J|}}.$$

Et en particulier, on a

$$\mathbb{P}(D_j = \varepsilon_j) = \frac{1}{2}.$$

D'où

$$\mathbb{P}\left(\bigcap_{j\in J} \{D_j = \varepsilon_j\} \right) = \prod_{j\in J} \mathbb{P}(D_j = \varepsilon_j).$$

Comme n et J sont arbitraires, cela prouve exactement que les D_j forment une suite de variables aléatoires indépendantes de loi $b(1/2)$.

- On commence par remarquer que

$$R_n = 2^n id - \sum_{j=1}^{n} 2^{n-j} D_j.$$

Soit f une fonction **positive mesurable** (pour la tribu borélienne), alors pour tout ε^n, on a

$$\begin{aligned}
E\left[f(R_n)\prod_{j=1}^{n}\mathbb{1}_{\{D_j=\varepsilon_j\}}\right] &= E\left[f(2^n id - \sum_{j=1}^{n} 2^{n-j}\varepsilon_j)\prod_{j=1}^{n}\mathbb{1}_{\{D_j=\varepsilon_j\}}\right] \\
&= \int_{\mathbb{R}} \mathbb{1}_{I^n_{\varepsilon^n}}(x) f\left(2^n x - \sum_{j=1}^{n} 2^{n-j}\varepsilon_j\right) \mathrm{d}\mathbb{P}(x) \\
&= \int_{\mathbb{R}} \mathbb{1}_{I^n_{\varepsilon^n}}\left(\frac{1}{2^n}y + \sum_{j=1}^{n}\frac{\varepsilon_j}{2^j}\right) f(y)\frac{1}{2^n}\mathrm{d}\mathbb{P}(y) \text{ (avec } y = 2^n x - \sum_{j=1}^{n} 2^{n-j}\varepsilon_j) \\
&= \frac{1}{2^n}\int_{\mathbb{R}} \mathbb{1}_{[0,1[}(y) f(y)\mathrm{d}\mathbb{P}(y) \\
&= \mathbb{P}\left(\bigcap_{j=1}^{n}\{D_j=\varepsilon_j\}\right)\int_{\mathbb{R}} \mathbb{1}_{[0,1[}(y) f(y)\mathrm{d}\mathbb{P}(y)
\end{aligned}$$

En sommant sur les ε^n, on obtient

$$E[f(R_n)] = \int_{\mathbb{R}} \mathbb{1}_{[0,1[}(y) f(y)\mathrm{d}\mathbb{P}(y).$$

On en déduit donc que R_n suit une loi uniforme.
Puis on peut répéter la même opération qu'auparavant en sommant sur $j \in J^c$ et on obtient

$$E\left[f(R_n)\prod_{j\in J}\mathbb{1}_{\{D_j=\varepsilon_j\}}\right] = \prod_{j\in J}\mathbb{P}(D_j=\varepsilon_j)E[f(R_n)].$$

Cela nous donne l'indépendance de R_n avec D_1, ..., D_n. □

Attention, les R_n ne sont pas indépendantes ! Sinon $D_n = 2R_{n-1} - R_n$ admettrait une densité, ce qui est faux.

Lemme.

Soit Y une variable aléatoire de loi uniforme sur $[0,1[$, F une fonction de répartition et G la fonction - appelée pseudo-inverse de F - définie par

$$\forall t \in \mathbb{R}, G(t) = \inf\{x \in \mathbb{R}, F(x) \geqslant t\}.$$

Alors $G(Y)$ a pour fonction de répartition F.

Démonstration. On a l'équivalence

$$F(x) \geqslant t \Leftrightarrow x \geqslant G(t).$$

Donc comme Y suit une loi uniforme sur $[0,1[$, on a

$$\mathbb{P}(G(Y) \leqslant x) = \mathbb{P}(Y \leqslant F(x)) = F(x).$$

□

Corollaire.

Soit $(\mu_j)_j$ une suite de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors il existe une suite $(X_j)_j$ de variables aléatoires **indépendantes** définies sur $([0,1[, \mathcal{B}([0,1[), \mathbb{P})$ et telle que X_j suive la loi μ_j.

Démonstration. On va créer une suite de variables aléatoires indépendantes $(Y_j)_j$ de lois uniformes sur $[0,1[$. Ainsi en utilisant le lemme, on aura terminé.

• On reprend les notations précédentes.
En utilisant la bijection $(\mathbb{N}^*)^2 \simeq \mathbb{N}^*$, [1]on peut partitionner $\mathbb{N}^*$ en union d'ensembles infinis

$$\mathbb{N}^* = \bigsqcup_{j\in\mathbb{N}^*} N_j,$$

et on note φ_j l'extractrice donnant les éléments de N_j dans l'ordre croissant.
On pose ensuite

$$Y_j = \sum_{k=1}^{\infty} \frac{1}{2^k} D_{\varphi_j(k)}.$$

• Par **le lemme des coalitions**, les Y_j sont indépendantes car ce sont des fonctions de différents D_k (et les D_k sont indépendantes).

• Montrons que Y_j suit une loi uniforme sur $[0,1[$.
On pose $Y_{j,n} = \sum_{k=1}^{n} \frac{1}{2^k} D_{\varphi_j(k)}$. Comme les D_k sont iid, $Y_{j,n}$ suit la même loi que

$$Z_n = \sum_{k=1}^{n} \frac{1}{2^k} D_k.$$

La suite Z_n tend presque surement vers Z la fonction identité.
De même, Y_j est la limite presque sure des $Y_{j,n}$. On a donc en particulier la convergence en loi, d'où

$$\mathbb{P}(Y_j \leqslant y) = \lim_n \mathbb{P}(Y_{j,n} \leqslant y) = \lim_n \mathbb{P}(Z_n \leqslant y) = \mathbb{P}(Z \leqslant y) = y.$$

Donc Y_j suit une loi uniforme sur $[0,1[$. □

Remarques : • Pour la leçon 249, on peut utiliser les Y_j pour avoir une suite de variables aléatoires iid de loi $b(p)$. Il suffit de poser $X_j = \mathbb{1}_{Y_j \leqslant p}$.
• La fonction de répartition d'une loi exponentielle est

$$F(x) = (1 - \exp(-\lambda x))\mathbb{1}_{[}0, \infty[.$$

Sa pseudo-inverse est alors

$$G(u) = -\lambda^{-1}\log(1-u).$$

De même, pour une loi de Cauchy, on a

$$G(u) = \tan(\pi(u - 1/2)).$$

• Pour les lois discrètes, il faut être plus malin. Si $\mu = \sum p_k \delta_{x_k}$, on se donne U une variable aléatoire uniforme sur $[0,1]$. On pose aussi la suite $(s_k)_k$ définie par $s_0 = p_0$ et $s_{k+1} = s_k + p_{k+1}$.
Alors la variable aléatoire X qui vaut x_k lorsque $U \in]s_{k-1}, s_k]$ suit la loi μ.
• Une utilité d'avoir une infinité de lois uniformes indépendantes est de pouvoir simuler une loi de Poisson de la manière suivante

$$N = \max\{n \in \mathbb{N}^*, U_1 ... U_n > \exp(-\lambda)\} \sim \mathcal{P}(\lambda).$$

1. On peut prendre l'application $(j,k) \mapsto k + \frac{(j+k-2)(j+k-1)}{2}$ par exemple.