249 - Suites de variables de Bernoulli indépendantes

Références [O1] Ouvrard, Probabilités 1.
[O2] Ouvrard, Probabilités 2.
[L] Emmanuel Lesigne, Pile ou face, Ellipses, 2001.

Cadre $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

I - Construction de suites de variables aléatoires de Bernoulli indépendantes

[O1] p.68

Définition [Variable de Bernoulli]
On appelle variable de Bernoulli, toute variable aléatoire X à valeurs entières dont la loi est donnée par $P(X=1)=p$ et $P(X=0)=1-p:=q$, avec $p\in[0,1]$.

Remarques * En général $0<p<1$.
* On parle d'une variable de Bernoulli de paramètre p et on note $X\sim p\delta_1+(1-p)\delta_0 \sim \mathcal{B}(1,p)$.

[O2] p.53 à 63.

Exemple [Modèle du pile ou face, suite finie de variables de Bernoulli]
. On peut construire un modèle probabiliste décrivant un jeu de pile ou face en n coups, grâce aux variables de Bernoulli.
. Mathématiquement cela revient à construire $(\Omega, \mathcal{A}, P)$ un espace probabilisé et n variables aléatoires indépendantes $(X_i)_{i=1..n}$ de même loi uniforme sur $[0,1[$ définies sur cet espace : on peut prendre $\Omega=\{0,1\}^n$ muni de la probabilité uniforme et pour X_i la projection sur le i-ème facteur : $(x_1,..,x_n)\mapsto x_i$

Problème [suites infinies]
Le problème analogue dans le cas d'une suite infinie de lancers devrait conduire à considérer $\Omega=\{0,1\}^{\mathbb{N}^*}$. Toutefois, l'existence sur cet espace d'une mesure de probabilité dont pour chaque i, la i-ème marginale soit la loi uniforme sur $[0,1[$ n'est pas évidente.

Résolution On se place sur $\Omega=[0,1[$ muni de la mesure de Lebesgue et $\forall x\in[0,1[$, on lui associe son développement dyadique $\in\{0,1\}^{\mathbb{N}^*}$.
Plus précisément :

Définition [Développement dyadique]. Pour $x\in[0,1[$, on définit les suites $D_n(x)$ et $R_n(x)$ par : $R_0(x)=x$ et $\forall n\in\mathbb{N}^*$, $D_n(x)=[2R_{n-1}(x)]$ et $R_n(x)=2R_{n-1}(x)-D_n(x)$.
$D_n(x)$ est le développement dyadique de x.

Remarques * $D_n(x)\in\{0,1\}$ et $R_n(x)\in[0,1[$.
* Le développement dyadique d'un réel n'est pas unique, mais la suite $(D_n(x))$ ici définit un unique développement $\forall x\in[0,1[$.

Proposition [Existence de suites de variables aléatoires de Bernoulli]. Soit $([0,1[, \mathcal{B}_{[0,1[}, \lambda)$ l'espace probabilisé où λ est la restriction de la mesure de Lebesgue à $[0,1[$. Alors sur cet espace, la suite $(D_n)_{n\in\mathbb{N}^*}$ est une suite de variables indépendantes de même loi de Bernoulli $\mathcal{B}(1,\frac{1}{2})$.

Corollaire [Existence de suites de variables aléatoires réelles de loi donnée]. Soit $(\mu_j)_{j\in\mathbb{N}^*}$ une suite de probabilités sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Alors il existe une suite de variables aléatoires réelles $(X_j)_{j\in\mathbb{N}^*}$ définies sur $([0,1[, \mathcal{B}_{[0,1]}, \lambda)$, indépendantes et telle que, $\forall j\in\mathbb{N}^*$, X_j soit de loi μ_j.

Remarque Ce corollaire se démontre via étapes :
- on prouve l'existence d'une suite de variables aléatoires indépendantes de même loi uniforme sur $[0,1[$, notée $(Y_j)_{j\in\mathbb{N}^*}$;

- On pose F_j la fonction de répartition de la probabilité μ_j, ie $\forall x \in \mathbb{R}, F_j(x) = \mu_j(]-\infty, x])$, on pose G_j sa pseudo-inverse, ie $\forall t \in \mathbb{R}$,

$$G_j(t) = \inf(x \,;\, F_j(x) \geq t)$$

et si $X_j = G_j(Y_j)$, on montre que la loi de X_j est μ_j.

Remarque [Simulation numérique] Grâce à la méthode de la pseudo-inverse, on peut simuler informatiquement toutes les suites de variables aléatoires réelles à partir d'un générateur de nombres pseudo-aléatoires (qui simule une loi uniforme sur $[0,1[$).

Remarque [Réciproque de la Proposition] À partir d'une suite de variable de Bernoulli de loi $\mathcal{B}(1,\frac{1}{2})$, il existe une construction qui permet de retrouver la mesure de Lebesgue sur $[0,1[$. Et si on considère une suite de Bernoulli de loi $\mathcal{B}(1,p)$, cette même construction nous fournit une mesure sur $[0,1[$ étrangère à la mesure de Lebesgue.

II - Construction de variables aléatoires à partir de variables aléatoires de Bernoulli.

[O 1] Chap. 7.

Proposition [Loi Binomiale] Soit $(X_i)_{1\leq i\leq n}$ une suite de variables de Bernoulli indépendantes, de paramètre p. Alors $S_n = \sum_{i=1}^{n} X_i$ est de loi Binomiale $\mathcal{B}(n,p)$

Proposition [Loi Géométrique] Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables de Bernoulli indépendantes, de paramètre p. On pose $A_n = (X_n = 1)$, $\forall n \in \mathbb{N}$. Alors la variable N définie $\forall \omega \in \Omega$ par $N(\omega) = \inf(n \in \mathbb{N} \,;\, \omega \in A_n)$ est de loi Géométrique sur $\mathbb{N}$, notée $\mathcal{G}_{\mathbb{N}}(p)$.

. Grâce au théorème qui suit, nous pouvons obtenir une approximation de la loi Binomiale $\mathcal{B}(n,p)$ lorsque n est "grand" et p "petit".

Théorème [de Poisson] Soit $(p_n)_{n\in\mathbb{N}}$ une suite de réels de $]0,1[$ telle que $\lim_{n\to\infty} n\, p_n = \lambda$ (où $\lambda > 0$). On considère, $\forall n \in \mathbb{N}$, une variable aléatoire S_n de loi $\mathcal{B}(n,p_n)$. Alors $\forall k \in \mathbb{N}$, la suite de terme général $P(S_n = k)$ est convergente et on a :

$$\lim_{n\to\infty} P(S_n = k) = \exp(-\lambda)\, \frac{\lambda^k}{k!}$$

. Grâce au théorème Central-Limite, nous pouvons approximer une gaussienne par des variables de Bernoulli. En effet :

Théorème [Central-Limite] Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires non constantes, de même loi, indépendantes, admettant un moment d'ordre 2, de moyenne $E(X_1)$ et d'écart-type σ_{X_1}. On pose $\forall n \in \mathbb{N}$, $S_n = \dfrac{\sum_{j=1}^{n} X_j - n E(X_1)}{\sqrt{n}\,\sigma_{X_1}}$.

Alors $\forall (a,b) \in \mathbb{R}^2$ tels que $a < b$, on a :

$$P(a < S_n \leq b) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx.$$

Corollaire [Théorème de Moivre-Laplace] Soit p tel que $0 < p < 1$. Soit S_n une variable de loi $\mathcal{B}(n,p)$. Soit $\tilde{S}_n = \dfrac{S_n - E(S_n)}{\sigma(S_n)}$ la variable aléatoire centrée réduite associée à S_n. Alors :

$$\sup_{-\infty \leq a < b \leq +\infty} \left| P(a < \tilde{S}_n \leq b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx \right| \xrightarrow[n\to\infty]{} 0 \,;$$

ie la loi de $\tilde{S}_n$ peut donc être approchée quand n est "grand" par la loi de Gauss centrée réduite $\mathcal{N}(0,1)$.

III - Applications

1) En statistique

. Grâce à la loi faible des grands nombres, on peut approcher le paramètre p d'une suite de variables de Bernoulli $\mathcal{B}(1,p)$ avec p inconnu.

[O 1] Chap 7.

Théorème [Loi faible des grands nombres]
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et de même loi, telle que X_1 admet un moment d'ordre 2. Alors la suite de variables aléatoires $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ converge en probabilité vers $m = E(X_1)$.

Corollaire [Théorème de Bernoulli] Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements indépendants de même probabilité p. Alors la suite des variables aléatoires $\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{A_j}$ converge en probabilité vers p.

2) En théorie des martingales

[O2] p.394 à 398.

Exemple [Ruine du Joueur] [Dvlpt 1]
- Un joueur joue à pile ou face avec une pièce non nécessairement équilibrée. Il reçoit 1€ de la banque s'il obtient pile et en donne un s'il obtient face. Le joueur joue jusqu'à sa ruine ou celle de la banque.
- Grâce à la théorie des martingales et à une suite de variables de Bernoulli indépendantes de même loi $p\delta_1 + (1-p)\delta_{-1}$, on déterminera :
  - la probabilité que le jeu s'arrête ;
  - la probabilité que le joueur gagne ;
  - le temps moyen d'arrêt du jeu.

3) En chaînes de Markov

[O2] p.422

- Grâce à la partie I, on sait qu'à partir d'une suite de variables de Bernoulli indépendantes, on peut construire la loi uniforme sur $[0,1[$ et avec cette loi uniforme, on peut construire des chaînes de Markov. En effet :

Proposition [Construction de chaînes de Markov]
Soit $g : E\times\mathbb{R} \to \mathbb{R}$ mesurable avec E un ensemble dénombrable muni de la tribu de ses parties. On considère une famille de variables aléatoires indépendantes, avec l'une de ces variables X_0 à valeurs dans E et les autres forment une suite de variables aléatoires réelles $(U_n)_{n\in\mathbb{N}^*}$ de même loi uniforme sur $[0,1[$. On construit $(X_n)_{n\in\mathbb{N}}$ par $X_{n+1} = g(X_n, U_{n+1})$.
Alors $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov relativement à la filtration $(\mathcal{A}_n)_{n\in\mathbb{N}}$ où $\mathcal{A}_0 = \sigma(X_0)$ et $\mathcal{A}_n = \sigma(X_0, U_1, \dots U_n)$ pour $n \geq 1$.

4) Estimation des grands écarts [Dvlpt 2]

[L] p.16 à 20.

- On s'intéresse ici à la vitesse de convergence dans la loi faible des grands nombres pour une suite de variables de Bernoulli.

Proposition Soit $p\in]0,1[$. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables de Bernoulli indépendantes de même loi $\mathcal{B}(1,p)$. On pose $S_n = \sum_{i=1}^{n} X_j \sim \mathcal{B}(n,p)$ pour $n\geq 1$.
Pour $\varepsilon \in]0, 1-p[$, on pose :
$h_+(\varepsilon) = (p+\varepsilon)\ln\left(\frac{p+\varepsilon}{p}\right) + (1-p-\varepsilon)\ln\left(\frac{1-p-\varepsilon}{1-p}\right)$.
Alors :
- $h_+(\varepsilon) > 0$;
- $\forall n\geq 1,\ P\left(\frac{S_n}{n} \geq p+\varepsilon\right) \leq e^{-n h_+(\varepsilon)}$;
- cette inégalité est optimale, ie $\frac{1}{n}\ln\left(P\left(\frac{S_n}{n}\geq p+\varepsilon\right)\right) \xrightarrow[n\to\infty]{} -h_+(\varepsilon)$.

Exemple [Utilisation de la Proposition]
Si on lance 1000 fois une pièce équilibrée, la probabilité que le nombre de piles soit supérieur ou égale à 600 est inférieure à 10^{-9}, ie ici $n = 1000$, $p = 0{,}5$ et $\varepsilon = 0{,}1$.