Leçon 251: Indépendance d'évènements et de variables aléatoires.

Antoine Barbé
Université de Rennes 1

Clément Guérin
ENS Cachan

11 avril 2012

Table des matières

Dans l'ensemble de la leçon $(\Omega, \mathcal{A}\ , P)$ est un espace probabilisé.

1 [P], Définitions et critères d'indépendance. [BAR]

1.1 Définitions.

Définition 1 *: (i) Une famille quelconque d'évènements* $(A_i) \in (\mathcal{A})^{\mathcal{I}}$ *est mutuellement indépendante si :* $\forall J \subset I \quad fini \qquad P(\bigcap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$.

Exemple : Jet de deux dés (Bleu, Rouge)
Soient A, B, C trois évènements représentant respectivement le dé rouge est impaire, le dé bleu est impair, la somme des deux dés est impair.
$P(A) = P(B) = P(C) = 1/2$ et $P(A \bigcap B) = P(A \bigcap C) = P(B \bigcap C) = 1/4$, mais $P(A \bigcap B \bigcap C) = \emptyset \neq 1/8$.
Ainsi, A, B, et C sont deux à deux indépendants mais pas mutuellement indépendants.

Définition 2 *:*
(ii) Une famille de sous-tribus $(A_i) \in P(\mathcal{A})^{\mathcal{I}}$ *est mutuellement indépendante si* $\forall A_i \in \mathcal{A}_\rangle$, $(A_i) \in \mathcal{A}^{\mathcal{I}}$ *est mutuellement indépendante.*
(iii) Soit (X_i) *famille de variables aléatoires de* $(\Omega, \mathcal{A}\ , P)$ *dans* $(E, \mathcal{B})$ *est mutuellement indépendante si la famille des tribus engendrées par les* X_i *est mutuellement indépendante, ie :*
$\forall J \subset I \quad fini, \quad \forall B_j \in \mathcal{B}\ P(X_j \in B_j | j \in J) := P(\bigcap \{X_j \in B_j\}) = \prod_{j\in J} P(X_j \in B_j)$.

Exemple : jet de deux dés.
Soient X_a, X_b deux variables aléatoires donnant respectivement le résultat du dérouge et celui du dé bleu. Ces deux variables sont indépendantes.

Remarque : L'indépendance de tribus est difficile à prouver en général, pour la prouver en pratique on la résultat suivant :

Propriété 1 *: Si* C_1 *et* C_2 *sont deux algèbre indépendantes (même définition que pour les tribus) alors* $\sigma(C_1)$ *et* $\sigma(C_2)$ *sont indépendantes.*

1.2 Indépendance de variables aléatoires et corrélation.

Propriété 2 *: Soit* $(X_1, ..., X_d)$ *une famille de variables aléatoires indépendantes réelles, alors* $p^{(x_1,...,x_d)}$ *la loi du vecteur aléatoire sur* $(\mathbb{R}^d, B(\mathbb{R}^d))$ *est égale au produit des lois marginales* $p^{X_1} \otimes ... \otimes p^{X_d}$.
Réciproquement si $p^{(x_1,...,x_d)} = p^{X_1} \otimes ... \otimes p^{X_d}$ *alors* $(X_1, ..., X_d)$ *est indépendantes.*

Exemple :Dans $\mathbb{R}^2$

		X_1		
	0	-1	0	1
X_2	-1	0	1/4	0
	0	1/4	0	1/4
	1	0	1/4	0

On pose $X = (X_1, X_2)$, on a : $p^X = \frac{1}{4}\delta_{(-1,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(0,-1)}$, alors $P(X = (0,0)) = 0 \neq \frac{1}{4} = P(X_1 = 0)P(X_2 = 0)$, ainsi X_1 et X_2 ne sont pas indépendantes.

Exemple : Indépendance non intuitive : Loi de γ et β [DÉVELOPPEMENT][OUV2] (p74).

Exemple : La loi uniforme sur $[0,1]^2$ a pour lois marginales les lois uniformes sur $[0,1]$.

Corollaire 1 *: Soit $(X_i)_{i\in I}$ suite de variables aléatoires réelles indépendantes si et seulement si $\forall J \subset I$ fini, $\forall \phi_j$ fonction borélienne telle que $\phi_j(X_j)$ intégrables, on ait :*

$$E(\prod \phi_j(X_j)) = \prod (E(\phi_j(X_j))).$$

Définition 3 *: Deux variables aléatoires X et Y dans $L^2(\Omega, \mathcal{A}, P)$ sont non corrélées si :*
$E(XY) = E(X)E(Y) \Leftrightarrow E((X - E(X))(Y - E(Y))) = Cov(X,Y) = 0 \Leftrightarrow X - E(X)$ et $Y - E(Y)$ sont orthogonales.

Remarque : D'après ce qui précède on a X, Y indépendantes implique X, Y non corrélées.

Contre-exemple : $Y = X^2$, $X \sim \mathcal{N}(0,1)$, $E(X) = 0$ et $E(XY) = 0$, ainsi $E(X)E(Y) = E(XY)$ mais X et Y ne sont pas indépendantes :
$P(|X| \leq 1, Y \geq 1) = 0 \neq P(|X| \leq 1)P(Y \geq 1)$.

1.3 Indépendance, fonctions caractéristiques et de répartitions.

Définition 4 *: $\varphi_X : \mathbb{R}^d \to \mathbb{R}$, $t \mapsto E(e^{i<t,k>})$ où X est une variable aléatoire réelle d-dimensionnelle.*

Propriété 3 *: Une famille $(X_1, ..., X_n)$ de variables aléatoires est une famille indépendante si et seulement si : $\forall (t_1, ..., t_n) \in \mathbb{R}^n$, $\varphi_{(X_1,..,X_n)}(t_1, ..., t_n) = \prod_{i=1}^n \varphi_{X_i}(t_i)$.*

Propriété 4 *: Soient X_1 et X_2 deux variables aléatoires, elles sont indépendantes si et seulement si $F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$.*

1.4 Corrélation et vecteur gaussien.

Définition 5 *Un vecteur $X = (x_1, ..., x_d) : (\Omega, \mathcal{A}\ , P) \to (\mathbb{R}^d, B(\mathbb{R}^d))$ est dit gaussien si $\forall \alpha \in \mathbb{R}$, $\sum_{i=1}^{d} \alpha_i X_i$ est une variable aléatoire gaussienne.*
Dans ce cas, $\sum \alpha_i X_i \sim \mathcal{N}\ (\sum \alpha_i E(X_i), \sum \alpha_i \alpha_j Cov(X_i, X_j))$.

Propriété 5 *: Un vecteur gaussien est entièrement déterminé par $m = E(X_i)$ et $\Gamma = Cov(X_i, X_j)$.*

Théorème 1.1 *: Soit X un vecteur gaussien de $\mathbb{R}^d$ de matrice de covariance Γ. Γ est diagonale si et seulement si $(X_1, ..., X_d)$ est indépendante.*

Remarque : La condition Γ diagonale est évidemment nécessaire (vrai dans le cas général), la réciproque fausse dans le cas général est vrai dans ce cas, néanmoins la condition vecteur gaussien est une condition forte :

Contre-exemple : Il ne suffit pas, pour que X soit gaussien que ses marginales soient gaussiennes :
$Z \sim \mathcal{N}\ (0, 1)$, ϵ de Rademacher : $\epsilon = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$.
Z et ϵ sont indépendants, alors $(Z, \epsilon Z)$ n'est pas gaussien mais de marginales gaussiennes.

2 [P], Nombre fini de variables aléatoires réelles indépendante

2.1 Somme de variables aléatoires indépendantes. [BAR]

Propriété 6 *: Soient X et Y deux variables aléatoires réelles indépendantes, la loi de la somme $X + Y$ est donnée par $p^X * p^Y$ où :*

Définition 6 *: $\forall \phi$ borélienne bornée de $\mathbb{R}$ dans $\mathbb{R}$:*
$\int_{\mathbb{R}} \phi d(p^X * p^Y) = \int_{\mathbb{R}}(\int_{\mathbb{R}} \phi(x + y)dp^Y(y))dp^X(x) = \int_{\mathbb{R}}(\int_{\mathbb{R}} \phi(x + y)dp^X(x))dp^Y(y)$.

Propriété 7 *:*
- $P * \delta_0 = P$
- $P * Q = Q * P$
- $(P * Q) * R = P * (Q * R)$
- $P * (\lambda Q + (1 - \lambda)R) = \lambda(P * Q) + (1 - \lambda)P * R$

Propriété 8 *Soient X et Y deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{A}\ , P)$, la fonction caractéristique de leur somme est donnée par :*
$\forall t \in \mathbb{R}$, $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$

Exemple :
- Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ alors $(X + Y) \sim \mathcal{P}(\lambda + \mu)$. $(\varphi_X(t) = e^{\lambda(e^{it}-1)})$

- Soient X et Y deux variables aléatoires indépendantes à densité f et g alors $(X+Y)$ a pour densité $f * g$.

.

Remarque : $\varphi_{X+Y} = \varphi_X \varphi_Y \nRightarrow X$ et Y indépendantes. [OUV2]
Pour le contre-exemple $Z = (X, Y)$ qui a pour densité $f(x, y) = \dfrac{1}{4} 1_C(x, y)[1+xy(x^2-y^2)]$ où $C = [-1, 1]^2$.

2.2 Retour sur les variables aléatoires gaussiennes. [OUV2](exo 13.3 et 13.4)

On donne dans cette partie deux caractérisations de lois gaussiennes.

$\rightarrow$ Caractérisation des lois gaussiennes sur $\mathbb{R}$:
X et Y deux variables aléatoires réelles admettant un moment d'ordre 2 de même loi μ telle que $E(X) = 0$, $E(X^2) = \sigma^2$ et X et Y indépendantes.
$\mu \sim \mathcal{N}\ (0, \sigma^2)$ si et seulement si $\dfrac{X+Y}{\sqrt{\epsilon}}$

$\rightarrow$ Caractérisation des variables aléatoires gaussiennes : Théorème de Bernstein.
X et Y deux variables aléatoires réelles indépendantes telles que $(X + Y)$ et $(X - Y)$ sont indépendantes, alors : X et Y sont des gaussiennes.

2.3 Lois infiniment divisibles. [COT](p135)

Définition 7 *:*
- *Soit $\mathcal{R}$ une loi de probabilité sur $(\mathbb{R}, B(\mathbb{R}))$ est infiniment divisible si $\forall n \in \mathbb{N}^*$, $\exists Q_n$ tel que $\mathcal{R} = Q_n^{*n}$*
- *De manière équivalente une variable aléatoire réelle X est infiniment divisible si $\forall n \in \mathbb{N}^*$, $\exists X_{1,n} ... X_{n,n}$ iid tels que $X \sim X_{1,n} + ... + X_{n,n}$.*

Propriété 9 *: X est infiniment divisible si et seulement si p_X l'est.*

Propriété 10 *: X est infiniment divisible si et seulement si $\forall n \in \mathbb{N}^*$, $\varphi_X = (\varphi_n)^n$ où φ_n est une fonction caractéristique.*

Exemple :
- $X \sim \mathcal{N}\ (\mu, \sigma^2) \longrightarrow X_{k,n} \sim \mathcal{N}\ (\dfrac{\mu}{n}, \dfrac{\sigma^2}{n})$
- $X \sim \mathcal{P}(\lambda) \longrightarrow X_{k,n} \sim \mathcal{P}\ (\dfrac{\lambda}{n})$
- $X \sim \gamma(a, \lambda) \longrightarrow X_{k,n} \sim \gamma(\dfrac{a}{n}, \lambda)$
- $X \sim \mathcal{C}\ (c) \longrightarrow X_{k,n} \sim \mathcal{C}\ (\dfrac{c}{n})$

Le théorème suivant est plus compliqué à démontrer, néanmoins il fournit une formule générale pour la fonction caractéristique d'une variable aléatoire infiniment divisible.

Théorème 2.1 *: X est infiniment divisible si et seulement si sa fonction caractéristique φ est de la forme :*

$$\varphi(u) = e^{i\beta u - \dfrac{\sigma^2 u^2}{2}} exp(\int (e^{iux} - 1 - \frac{iux}{1+x^2})(\frac{1+x^2}{x^2})\gamma(dx)$$

où $(\beta, \sigma) \in \mathbb{R}^{\not{\vDash}}$ et γ mesure finie ne changeant pas l'origine.

3 [P], Convergence de suites et indépendance.

3.1 Convergence en loi de somme de variables aléatoires iid. [COT]

Propriété 11 *: $X_n \xrightarrow{\mathcal{L}} X$ où $(X_n), X$ sont des variables aléatoires. Alors (X_n) est tendue, ie $\forall \epsilon > 0, \exists$ compact K tel que $\forall n \in \mathbb{N}$, $P(X_n \in K) \geq 1 - \epsilon$.*

Théorème 3.1 *: Helly : Soit (F_n) une suite de fonctions de répartitions, $\exists F$ croissante continue à gauche tel que $0 \leq F \leq 1$ et une sous-suite F_{n_k} tel que :*

$$F_{n_k}(t) \to F(t) \text{ en tout point } t \text{ de continuité de } F.$$

Théorème 3.2 *: Prohorov : Soit ϕ une famille de probabilité de $(\mathbb{R}, B(\mathbb{R}))$, les propriétés suivantes sont équivalentes :*

(i) $\forall \epsilon > 0$, $\exists K$ compact de $\mathbb{R}$ tel que : $\forall Q \in \phi$, $Q(K) \geq 1 - \epsilon$.

(ii) De toute suite (Q_n) d'éléments de ϕ, $\exists n_k$ et Q tels que : $Q_{n_k} \xrightarrow{\mathcal{L}} Q$.

Corollaire 2 *: Soient $m \in \mathbb{N}$ fixé et $\forall i \in \mathbb{N}$, $(X_{i,m})iid$, on pose $S_m = X_{1,m} + ... + X_{m,m}$ et on a $S_m \xrightarrow{\mathcal{L}} S$, alors S est infiniment divisible.*

Remarque : Toutes les convergences considérées par la suite seront dans ce cadre là.

3.2 Loi faible des grands nombres. [ZQ]

Propriété 12 *:*
- *Inégalité de Markov :*
$$X \in L^1,\ t > 0,\ P(|X| \geq t) \leq \frac{E(|X|)}{t}.$$
- *Inégalité de Tchebycheff :*
$$X \in L^2,\ t > 0,\ P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}.$$

Théorème 3.3 *: Loi faible des grands nombres :*
Soient $X \in L^2$, $(X_n)_{n\geq 1}$ *iid et* $S_n = \sum_{i=1}^n X_i$, *alors* $\forall t > 0$, $P(|\frac{S_n}{n} - E(X)| > t) \leq \frac{Var(X)}{nt^2}$.
En particulier $\frac{S_n}{n} \xrightarrow{P} E(X)$.

Dans le cadre de la loi faible des grands nombres, on peut démontrer le théorème de Weierstrass sur $[0,1]$ par les polynômes de Bernstein :

Théorème 3.4 *: [DÉVELOPPEMENT]*
soit $f : [0,1] \to \mathbb{C}$ *continue, alors* $||f - B_n(f)||_\infty \leq C\omega(\frac{1}{\sqrt{n}})$, *où* ω *module de continuité de* f. *De plus, c'est optimal.*

3.3 Lemme de Borel-Cantelli et loi forte des grands nombres. [BAR]

Définition 8 *: Soit* (T_n) *une famille de tribus, indépendante.* $\mathcal{A}_n = \sigma(T_n, T_{n+1})$, *on pose* $\mathcal{A}_\infty = \bigcap \mathcal{A}_n$ *est la tribu terminale.*

Théorème 3.5 *: Loi du* $0-1$ *:*
Si $\mathcal{A}_\infty$ *tribu terminale, alors* $\forall A \in \mathcal{A}_\infty$, $P(A) = 0$ *ou* 1.

Définition 9 *: Soit* (A_n) *une suite d'événements* $A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = \{A_n$ *a lieu une infinité de fois* $\}$. *On le note* A_n *i.s. (infiniment souvent), on le note aussi* $limsup A_n$.

Théorème 3.6 *: Lemme de Borel-Cantelli :*
(A_n) *suite d'événements*

i) $\sum P(A_n) < +\infty \Rightarrow P(A_n i.s.) = 0$.
ii) Si (A_n) *indépendants alors* $\sum P(A_n) = \infty \Rightarrow P(A_n i.s.) = 1$.

<u>Contre-exemple :</u>ii)indépendant : $\Omega = [0,1]$, $A_n =]0, \frac{1}{n}]$ alors $\sum P(A_n) = \infty$ et $P(A_n i.s.) = 0$.

Théorème 3.7 *: Loi forte des grands nombres :*
Soit (X_i) *suite iid telle que* $\forall i$*, et* $X_i \sim X$ *et* $S_n = \sum_{i=1}^{n} X_i$*, alors :*

$$E(|X|) < +\infty \textit{ si et seulement si } \frac{S_n}{n} \to E(X) \textit{ presque sûrement.}$$

<u>Contre-exemple :</u> Si $X_i \sim \mathcal{C}\ (C)$ alors $\frac{S_n}{n} \sim \mathcal{C}\ (C)$, donc $\frac{S_n}{n} \to X$ mais $E(|X|) = +\infty$.

<u>Exemple :</u> U identité sur $[0,1]$, $\omega \in [0,1]$, $U(\omega) = \sum 2^{-i} U_i(\omega)$ développement dyadique de $U(\omega)$. $U_i \sim B(1, \frac{1}{2})$ et (U_i) indépendants.
On a, par la loi forte des grands nombres, $P(\omega : lim_{n\to\infty} \frac{1}{n} \sum_{1\leq i\leq n} U_i(\omega) = \frac{1}{2}) = 1$. Ainsi presque tout ω dans $[0,1]$ admet en moyenne autant de 0 que de 1 dans son développement dyadique.

3.4 Théorème central-limite. [BAR]

Théorème 3.8 *: Central-limite :*
Soit X_i *iid tels que* $X_i \sim X$*, on pose* $S_n = \sum_{i=1}^{n} X_i$*, on a :*

i) si $E(X^2) < \infty$ *alors* $\frac{S_n - nE(X)}{\sqrt{nVar(X)}} \xrightarrow{\mathcal{L}} \mathcal{N}\ (0,1)$.

ii) si $\frac{S_n}{\sqrt{n}}$ *converge en loi, alors* $E(X) = 0$*,* $E(X^2) < \infty$ *et la loi limite est normale centrée de variance* $Var(X)$.

Applications :

- Critère empirique d'indépendance.
- Statistique.

<u>**BIBLIOGRAPHIE :**</u>

BAR : Barbe, Ledoux <u>Probabilité</u>.
OUV2 : Oubrad <u>Probabilités 2</u>.
Cot : Cottrel, ..., <u>Exercices de probabilités</u>.
ZQ : Zuily Queffelec <u>Analyse pour l'agrégations</u>.
Bonne Lecture !

4 Développements

4.1 Indépendance non-intuitive

Ce développement permet d'illustrer une méthode pour démontrer l'indépendance de variables aléatoires à densité U et T : montrer que la densité du couple s'écrit comme le produit des densités. L'énoncé est le suivant :

Propriété 13 *Soit X et Y indépendantes de loi $\gamma(a,p)$ et $\gamma(b,p)$ où a,b,$p > 0$.*
On pose $U = X + Y$ et $T = \frac{X}{X+Y}$.

Alors on a $U \perp T$. On peut de plus déterminer les lois de U et T.

On notera $C_c^+(\mathbb{R}^2)$ l'ensemble des fonctions continues positives à support compact, E l'espérance et pour toute variable aléatoire X à densité f_X sa densité.

Pour commencer il faut vérifier que T soit bien définie, en effet $X \perp Y$, X à densité et Y à densité impliquent que $X + Y$ est à densité et donc :

$$P(X + Y = 0) = 0$$

Ainsi T est bien définie (presque sûrement).

Soit $f \in C_c^+(\mathbb{R}^2)$, il s'agit d'écrire $E(f(U,T))$ sous la forme $\int_{\mathbb{R}^2} f(u,t)g(u,t)du\,dt$ où g est une fonction mesurable sur $(\mathbb{R})^2$ (qui ne dépend pas de f), g sera alors la densité de (U,T), puis $g(u,t) = a(u)b(t)$ où a et b sont des densités sur $\mathbb{R}$, on aura alors que U est de densité a, T de densité b et $U \perp T$.

$E(f(U,T)) = E(f(X+Y, \frac{X}{X+Y})) = \int_{\mathbb{R}^2} f(x+y, \frac{x}{x+y})f_X(x)f_Y(y)dx\,dy$

On utilise ici l'indépendance de X et Y et le théorème de transfert.

On explicite maintenant les densités de X et Y nulles sur pour tout $t \leq 0$:

$$f_X(x) = \frac{1}{\Gamma(a)p^a}e^{-px}x^{a-1}$$

$$f_Y(y) = \frac{1}{\Gamma(b)p^b}e^{-py}y^{b-1}$$

$E(f(U,T)) = \int_{(\mathbb{R}^{+*})^2} f(x+y, \frac{x}{x+y})\frac{1}{\Gamma(a)\Gamma(b)p^{a+b}}e^{-p(x+y)}x^{a-1}y^{b-1}dx\,dy$

Afin d'alléger les notations on écrira $C = \frac{1}{\Gamma(a)\Gamma(b)p^{a+b}}$.
Pour calculer l'intégrale, il va falloir utiliser un changement de variables.
Soit $g : (\mathbb{R}^{+*})^2 \to \mathbb{R}^{+*} \times (0,1)$ définie par :

$$g(x,y) = (x+y, \frac{x}{x+y})$$

Soit $h : \mathbb{R}^{+*} \times (0,1) \to (\mathbb{R}^{+*})^2$ définie par :

$$h(u,t) = (ut, u(1-t))$$

Alors g et h sont C^1 sur leur ensemble de définition et de plus :

$$g \circ h = Id_{\mathbb{R}^{+*}\times(0,1)}$$

$$h \circ h = Id_{(\mathbb{R}^{+*})^2}$$

donc h est un C^1 difféomorphisme.

Calculons sa matrice jacobienne :

$$Dh(u,t) = \begin{pmatrix} t & u \\ 1-t & -u \end{pmatrix}$$

Le déterminant de cette matrice est $-u$.

$E(f(U,T)) = C \int_{(\mathbb{R}^{+*})^2} f \circ g(x,y) e^{-p(x+y)} x^{a-1} y^{b-1} dx\, dy$

par changement de variable on obtient :
$E(f(U,T)) = C \int_{\mathbb{R}^{+*}\times(0,1)} f(u,t) |Dh(u,t)| e^{-pu} (ut)^{a-1} (u(1-t))^{b-1} du\, dt$
$E(f(U,T)) = C \int_{\mathbb{R}^{+*}\times(0,1)} f(u,t) e^{-pu} u^{a+b-1} t^{a-1} (1-t)^{b-1} du\, dt$

Par suite on obtient la densité de (U,T) :

$$f_{(U,T)}(u,t) = C 1_{\mathbb{R}^{+*}}(u) e^{-pu} u^{a+b-1} 1_{(0,1)}(t) t^{a-1} (1-t)^{b-1}$$

Remarquons qu'à ce stade on sait que $f_{(U,T)}$ est une densité donc $\int\int f_{(U,T)} = 1$ donc si l'on note :

$$a : u \to 1_{\mathbb{R}^{+*}}(u) e^{-pu} u^{a+b-1}$$

$$b : t \to 1_{(0,1)}(t) t^{a-1} (1-t)^{b-1}$$

on sait que a et b sont positives, de plus $\int a = \Gamma(a+b) p^{a+b}$ et $C = \frac{1}{\Gamma(a)\Gamma(b)p^{a+b}}$. Donc si l'on note :

$$\tilde{a} : u \to \frac{1}{\Gamma(a+b)p^{a+b}} a(u)$$

$$\tilde{b} : t \to \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} b(u)$$

on a $f_{(U,T)}(u,t) = \tilde{a}(u)\tilde{b}(t)$ où $\tilde{a}$ et $\tilde{b}$ sont des densités. On en déduit que U et T sont indépendantes et de plus :

$$f_U(u) = \frac{1}{\Gamma(a+b)p^{a+b}} 1_{\mathbb{R}^{+*}}(u) e^{-pu} u^{a+b-1}$$

$$f_T(t) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} 1_{(0,1)}(t) t^{a-1}(1-t)^{b-1}$$

Ce développement (qui est un exemple concret de démonstration d'indépendance) rentre tout à fait dans le cadre de la leçon et permet d'introduire une loi nommée β de paramètres a, b, p dont la densité est f_T. Ce développement est à la page 74 du livre "Probabilités" d'Ouvrard.

4.2 Théorème de Weierstrass par les polynômes de Bernstein

Dans ce développement on se propose de donner une approximation par des polynômes d'une fonction $f : [0,1] \to \mathbb{C}$ continue et d'évaluer cette approximation, il faut pour cela définir le module de continuité d'une fonction continue.

Définition 10 (module de continuité) *Soit $f : [0,1] \to \mathbb{C}$ continue, le module de continuité noté ω de f est défini par :*

$$\omega(h) = sup \{|f(x) - f(y)| \, |x, y \in [0,1], |x-y| < h\}, \forall h \geq 0$$

La fonction f étant continue sur un compact, f est uniformément continue, le module de continuité est bien défini. Il vérifie les propriétés suivantes faciles à démontrer :

Propriété 14 *Soit $f : [0,1] \to \mathbb{C}$ continue et ω son module de continuité alors :*
- *ω est une fonction croissante.*
- *ω est une fonction sous-additive : $\forall x, y$, $\omega(x+y) \leq \omega(x) + \omega(y)$.*
- *$\forall \lambda, h \geq 0$, $\omega(\lambda h) \leq (\lambda+1)\omega(h)$*

Les deux premiers points sont évidents avec la définition, pour le troisième point, on montre pour tout entier n la propriété suivante par récurence (avec le deuxième point) : $\omega(nh) \leq n\omega(h)$. Puis, en utilisant la partie entière de λ et avec la croissance on peut conclure.

On définit maintenant le polynôme qui va faire l'approximation :

Définition 11 (Polynômes de Bernoulli associés à une fonction) *Soit $f : [0,1] \to \mathbb{C}$ continue, $n \in \mathbb{N}$ le n-ième polynôme de Bernoulli associé à f est définie par :*

$$B_n(f)(x) = \sum_{k=0}^{n} (C_n^k) x^k (1-x)^{n-k} f(\frac{k}{n}), \forall x \in [0,1]$$

On peut maintenant énoncé le résultat à démontrer :

Propriété 15 *Soit $f : [0,1] \to \mathbb{C}$ continue, ω son module de continuité et $(B_n(f))$ ses polynômes de Bernoulli associés alors, il existe $C > 0$ tel que :*

$$||f - B_n(f)||_\infty \leq C\omega(\frac{1}{\sqrt{n}})$$

De plus on peut exhiber une fonction f_0 continue (ω_0 son module de continuité), $C_0 > 0$ telles que :

$$||f_0 - B_n(f_0)||_\infty \geq C_0\omega_0(\frac{1}{\sqrt{n}})$$

Autrement dit, l'estimation de l'approximation est optimale (l'ordre de grandeur, pas forcément pour ce qui est de la constante).

Toute la preuve du premier point est contenue dans l'interprétation probabiliste de $B_n(f)(x)$.

Soit $x \in [0,1]$, $n \in \mathbb{N}^*$, $(X_i)_{1\leq i\leq n}$ n variables indépendantes de même loi de Bernoulli de paramètre x et $S_n = \sum_{k=0}^n X_k$.

$$E(f(\frac{S_n}{n})) = \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} f(\frac{k}{n}) = B_n(f)(x)$$

Remarquons que la loi faible des grands nombres implique ici que $B_n(f)(x) \to_n f(x)$, mais l'approximation n'est pas estimée (et même pas uniforme en x).

$$\begin{array}{rcll}
|f(x) - B_n(f)(x)| & \leq & E(|f(x) - B_n(f)(x)|) & \\
& \leq & E(\omega(|x - \frac{S_n}{n}|)) & \textit{par définition de } \omega \\
& \leq & E((\sqrt{n}|x - \frac{S_n}{n}| + 1)\omega(\frac{1}{\sqrt{n}})) & \textit{par propriété de } \omega \\
& \leq & \omega(\frac{1}{\sqrt{n}})(1 + \sqrt{n}||x - \frac{S_n}{n}||_1) & \textit{car } ||g||_1 = E(|g|) \\
& \leq & \omega(\frac{1}{\sqrt{n}})(1 + \sqrt{n}||x - \frac{S_n}{n}||_2) & \textit{car } ||g||_1 \leq ||g||_2 \\
& \leq & \omega(\frac{1}{\sqrt{n}})(1 + \sqrt{n}\sqrt{\frac{x(1-x)}{n}}) & \\
& \leq & \omega(\frac{1}{\sqrt{n}})(1 + x(1-x)) & \\
& \leq & \frac{5}{4}\omega(\frac{1}{\sqrt{n}}) & \textit{car } x(1-x) \leq \frac{1}{4}
\end{array}$$

Le premier point est donc démontré.

Posons f la fonction continue de $[0,1]$ dans $\mathbb{C}$ définie par $f(x) = |x - \frac{1}{2}|$.

On note ω son module de continuité, alors en utilisant l'inégalité triangulaire on a : $\omega(h) \leq h$.

Ici, nous devons démontrer un lemme :

Lemme 1 (Inégalité de Khintchine) *Soit $(\epsilon_i)_{1\leq i\leq n}$, n variables indépendantes identiquement distribuées de même loi $\frac{1}{2}\delta_{-1}+\frac{1}{2}\delta_1$. On appelle de telles variables des variables de Rademacher.*

Soit $(a_j)_{1\leq i\leq n}$ des réels et $f=\sum_{j=1}^n a_j\epsilon_j$, alors on a :

$$||f||_1 \geq \frac{1}{e}||f||_2$$

Soit $1\leq i,j\leq n$ alors on a $E(\epsilon_i)=0$ et $E(\epsilon_i\epsilon_j)=\delta_{i,j}$ par indépendance.
Par suite, $||f||_2=\sum(a_j)^2$.

Remarquons que par homogénéité on peut supposer $||f||_2=1$ (i.e. $\sum(a_j)^2=1$).

L'idée est de démontrer ce résultat par dualité, en effet pour tout $g\in L^\infty$ on a $|E(fg)|\leq ||f||_1||g||_\infty$, d'où :

$$||f||_1 \geq \frac{|E(fg)|}{||g||_\infty}$$

Posons maintenant $g=\prod_{j=1}^n(1+ia_j\epsilon_j)$, c'est le produit de Riesz imaginaire.
presque pour tout ω, on a $|g(\omega)|=\prod_{j=1}^n\sqrt{1+(a_j)^2(\epsilon_j)^2}=\prod_{j=1}^n\sqrt{1+(a_j)^2}\leq\prod_{j=1}^n\sqrt{e^{(a_j)^2}}=\sqrt{e}$.
Par suite,

$$||g||_\infty \leq \sqrt{e}$$

Calculons pour tout $1\leq k\leq n$, $E(\epsilon_k g)$:
$E(\epsilon_k g)=E(\epsilon_k\prod_{j=1}^n(1+ia_j\epsilon_j))$.
$E(\epsilon_k g)=E(\epsilon_k(1+ia_k\epsilon_k)\prod_{j=1,j\neq k}^n(1+ia_j\epsilon_j))$.
$E(\epsilon_k g)=E(\epsilon_k(1+ia_k\epsilon_k))\prod_{j=1,j\neq k}^n E((1+ia_j\epsilon_j))$, par indépendance.
$E(\epsilon_k g)=ia_k\prod_{j=1,j\neq k}1$.
$E(\epsilon_k g)=ia_k$

Par linéarité $|E(fg)|=|\sum_{k=1}^n i(a_k)^2|=1$.
On a donc :

$$||f||_1 \geq \frac{|E(fg)|}{||g||_\infty} \geq \frac{1}{\sqrt{e}}$$

Cela termine la démonstration du lemme.

Revenons à f :
$||f-B_n(f)||_\infty \geq |f(\frac{1}{2})-B_n(f)(\frac{1}{2})|=|B_n(f)(\frac{1}{2})|$.

Il ne reste plus qu'à faire apparaître des variables de Rademacher ϵ_i indépendantes, puis d'appliquer l'inégalité de Khintchine :

$$|B_n(f)(\frac{1}{2})| = E(|\frac{S_n}{n} - \frac{1}{2}|) = \frac{1}{2n}E(|2S_n - n|)$$

Puis en remarquant que $2S_n - n = \sum_{i=1}^{n}(2X_i - 1)$, on obtient en posant $\epsilon_i = 2X_i - 1$:

$$|B_n(f)(\frac{1}{2})| = E(|\sum_{i=1}^{n}\epsilon_i|)$$

On est dans le cas du lemme (avec $a_i = 1$) en appliquant l'inégalité de Khintchine on obtient :

$$|B_n(f)(\frac{1}{2})| \geq \frac{1}{2n\sqrt{e}}||\sum_{i=1}^{n}\epsilon_i||_2 = \frac{1}{2n\sqrt{e}}\sqrt{Var(\sum_{i=1}^{n}\epsilon_i)}$$

Enfin avec $Var(\sum_{i=1}^{n}\epsilon_i) = \sum_{i=1}^{n}Var(\epsilon_i) = n$ par indépendance des ϵ_i et $Var(\epsilon_i) = 1$, on obtient :

$$|B_n(f)(\frac{1}{2})| \geq \frac{1}{2\sqrt{n}\sqrt{e}} \geq \frac{1}{2\sqrt{e}}\omega(\frac{1}{\sqrt{n}})$$

.

Ce développement est un peu long mais a le mérite de mélanger analyse et probabilités, l'utilisation de l'inégalité de Khintchine par exemple est trés appréciée. Remarquons que sa démonstration fait intervenir des calculs élémentaires (voir triviaux), on peut gagner du temps en ne donnant que les étapes importantes (ce lemme contenant une bonne partie du coeur de la preuve, il serait malvenu de l'admettre entièrement).

Ce développement est issu du livre "Analyse pour l'agrégation" de Zuilly et Queffélec.

5 Remarques complémentaires

Dans cette partie nous allons voir quelques éléments qui auraient pu figurer dans la leçon mais qui n'y étaient pas par manque de place ou de connaissance et expliquer quelques uns des choix de la leçon.

5.1 Définir l'indépendance

Il faut définir, dans une leçon sur l'indépendance, l'indépendance entre évenements, entre variables aléatoires et entre tribus.

De plus il faut pouvoir justifier la cohérence entre ces définitions (cela nécessite de pouvoir parler de tribu engendrée par une variable aléatoire).

La remarque sur la preuve effective d'indépendance entre tribus est nécessaire (il faut donc connaître des éléments un peu fondamentaux sur la théorie des probabilités : σ-algèbre...).

Enfin, il faut essayer de faire une liste la plus exhaustive possible des critères d'indépendance et tôt dans la leçon. Les critères d'indépendance doivent être connus, en cas de question, sans regarder son plan.

Pour l'essentiel, nous avons suivis le livre "probabilités" de Barbe et Ledoux.

5.2 Expression de $\frac{sin(t)}{t}$

Ici nous allons démontrer, de manière probabiliste, le théorème suivant :

Théorème 5.1 $\frac{sin(t)}{t} = \prod_{j\geq 1} cos(\frac{t}{2^j})$*, pour tout* $t \in \mathbb{R}$.

Avant de commencer la preuve de ce théorème rappelons ici que c'est une formule classique dont il existe des démonstrations d'analyse pur plus simple (voir sujet d'agrégation d'analyse de l'année 1996 par exemple). Il ne faudrait donc pas présenter ce développement comme une application à proprement parler mais plutôt comme une autre façon de voir les choses (ce qui, d'un point de vue pédagogique, est tout à fait défendable).

Remarquons que $\frac{sin(t)}{t} = \frac{e^{it}-e^{-it}}{t} = \int_{-1}^{1} e^{itx}dx = \phi_U(t)$, où U est la loi uniforme sur $[-1, 1]$.

Ainsi, on peut voir, dans notre identité à démontrer, une simple convergence de fonctions caractéristiques.

Notre plan de démonstration est donc clair, expliciter une suite de variables aléatoires (bien choisies) dont les fonctions de répartition convergent vers celle de la loi uniforme. On aura la convergence en loi et donc la convergence des fonctions caractéristiques.

Soit $n \in \mathbb{N}^*$, $(X_i)_{1\leq i\leq n}$, n variables indépendantes de même loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, (on retrouve ici des variables aléatoires de Rademacher). On pose $S_n = \sum_{i=1}^{n} \frac{1}{2^i} X_i$. Affirmons dés maintenant que ce sont les S_n qui vont être nos variables aléatoires bien choisies. Commençons par expliciter la loi des S_n :

Propriété 16 (Loi des S_n) S_n *a pour loi :* $\frac{1}{2^n}(\sum_{k=0}^{2^n-1} \delta_{\frac{2k+1}{2^n}-1})$

La démonstration se fait naturellement par récurence sur n. Le cas $n = 1$ est évident.

Soit donc $n \geq 1$ tel que l'on ait la loi de S_n, $k \in \{0...2^{n+1} - 1\}$, en remarquant que $S_{n+1} = S_n + \frac{1}{2^{n+1}} X_{n+1}$:

$P(S_{n+1} = \frac{2k+1}{2^{n+1}} - 1) = P(S_n = \frac{2k+1}{2^{n+1}} + \frac{1}{2^{n+1}} - 1, X_{n+1} = -1) + P(S_n = \frac{2k+1}{2^{n+1}} - \frac{1}{2^{n+1}} - 1, X_{n+1} = 1)$

Puis par indépendance on obtient :

$P(S_{n+1} = \frac{2k+1}{2^{n+1}}) = P(S_n = \frac{2(k+1)}{2^{n+1}} - 1)P(X_{n+1} = -1) + P(S_n = \frac{2k}{2^{n+1}} - 1)P(X_{n+1} = 1)$
$P(S_{n+1} = \frac{2k+1}{2^{n+1}}) = \frac{1}{2}P(S_n = \frac{(k+1)}{2^n} - 1) + \frac{1}{2}P(S_n = \frac{k}{2^n} - 1)$

Maintenant si k est pair alors $k+1 = 2\tilde{k}$ où $\tilde{k} \in \{0...2^n - 1\}$ et on obtient (en utilisant l'hypothèse de récurence) :

$$P(S_{n+1} = \frac{2k+1}{2^{n+1}}) = \frac{1}{2^{n+1}} + 0 = \frac{1}{2^{n+1}}$$

Et si k est impair alors $k = 2\tilde{k} + 1$ où $\tilde{k} \in \{0...2^n - 1\}$ et on obtient (en utilisant l'hypothèse de récurence) :

$$P(S_{n+1} = \frac{2k+1}{2^{n+1}}) = 0 + \frac{1}{2^{n+1}} = \frac{1}{2^{n+1}}$$

Dans tous les cas $P(S_{n+1} = \frac{2k+1}{2^{n+1}}) = \frac{1}{2^{n+1}}$, cela pour tout $k \in \{0...2^{n+1} - 1\}$.
Pour conclure il suffit de remarquer que $\sum_{k=0}^{2^{n+1}-1} P(S_{n+1} = \frac{2k+1}{2^{n+1}}) = 1$, cela nous dit que l'on a obtenu tous les points chargés par S_{n+1} et donc on a la propriété au rang $n+1$, puis on a la loi de S_n, pour tout entier n par récurence.

Soit $n \in \mathbb{N}^*$ alors si l'on note F_X la fonction de répartition de X :

$F_{S_n}(t) = 0$, si $t < \frac{1}{2^n} - 1$.
$F_{S_n}(t) = \frac{k+1}{2^n}$, si $t \in \left[\frac{2k+1}{2^n} - 1, \frac{2(k+1)+1}{2^n} - 1\right)$ où $k \in \{0...2^n - 1\}$.
$F_{S_n}(t) = 1$, si $t > 1 - \frac{1}{2^n}$.

D'autre part si l'on note F la fonction de répartition de la loi uniforme :

$F(t) = 0$, si $t < -1$.
$F(t) = \frac{t+1}{2}$, si $t \in [0, 1]$.
$F(t) = 1$, si $t > 1$.

Remarquons que si $t \leq -1$ on a, pour tout n, $F_{S_n}(t) = 0$ et donc $F_{S_n}(t) \to_n F(t)$.
Si $t \geq 1$ on a,pour tout n, $F_{S_n}(t) = 1$ et donc $F_{S_n}(t) \to_n F(t)$.

Soit donc $t \in (0, 1)$, il existe N tel que pour tout $n > N$ on ait $t \in \left(\frac{1}{2^n} - 1, 1 - \frac{1}{2^n}\right)$.

Dans ce ca là $F_{S_n}(t) = \frac{k+1}{2^n}$. Il suffit ensuite de remarquer que $-\frac{1}{2^n} < \frac{k+1}{2^n} - \frac{t-1}{2} < \frac{1}{2^n}$:

$$|F_{S_n}(t) - F(t)| < \frac{1}{2^n}$$

Ainsi pour tout $t \in (0,1)$ on a $F_{S_n}(t) \to_n F(t)$.

On a montré $F_{S_n} \to_n F$ simplement sur $\mathbb{R}$, d'après la théorie des probabilités, cela implique que $S_n \xrightarrow{L} U$, où U est la loi uniforme sur $[-1,1]$, puis d'aprés la même théorie des probabilités, on a la convergence simple des fonctions caractéristiques (notons ϕ_X, la fonction caractéristique de la variable aléatoire X).

Il ne reste plus qu'à calculer $\phi_{S_n} = \phi_{\sum_{j=1}^n \frac{1}{2^j} X_j} = \prod_{j=1}^n \phi_{\frac{1}{2^j} X_j}$, par indépendance.

Pour tout $t \in \mathbb{R}$, on a :

$$\phi_{\frac{1}{2^j} X_j}(t) = E(e^{\frac{1}{2^j} itX_j}) = \frac{1}{2}(e^{\frac{1}{2^j} it} + e^{\frac{1}{2^j} it}) = cos(\frac{t}{2^j})$$

Ainsi $\phi_{S_n}(t) = \prod_{j=1}^n cos(\frac{t}{2^j})$, d'après la convergence des fonctions caractéristiques on obtient le résultat annoncé.

Dans cette démonstration, on trouve plusieurs choses, indépendance, convergence, utilisation de variables de Rademacher, développement dyadique des réels entre -1 et 1 ... Le défaut majeur est le côté calculatoire (loi des S_n, fonctions de répartition), défaut qui peut être court-circuité en dessinant au tableau les premières fonctions de répartition, on voit clairement apparaître la fonction de répartition de la loi uniforme, une fois que l'on a dessiné cela et avec une argumentation intelligente, on peut sauter les calculs en expliquant avec les mains la convergence des fonctions de répartition.

Ce développement relativement connu des probabilistes n'est pas bien référencé, et l'on ne peut pas ici en donner une référence. Il peut aussi tomber en exercice.

5.3 Autour des théorèmes limites

Que se passe-t-il dans les théorèmes quand les variables aléatoires ne sont pas indépendantes ? Il faudrait préparer quelques contre-exemples, cela a sa place dans le cadre de la leçon car cela met en valeur l'importance cruciale de l'indépendance, de plus cette reflexion peut se retrouver dans la leçon sur les théorèmes limites.

Exemple 1 (Loi faible) *Si $X_i = X$, pour tout i avec X loi normale centrée de variance 1.*

Alors on a $S_n = X$ presque sûrement donc convergence en probabilité de S_n vers un loi normale (et pas un dirac en $0 = E(X)$).

Une autre chose que l'on peut essayer de contredire est l'intégrabilité dans la loi forte, par exemple :

Exemple 2 (Loi forte et loi de Cauchy) *Si (X_i) sont indépendantes de même loi de Cauchy de paramètre a, alors la loi de S_n est une loi de Cauchy de paramètre a (pour le démontrer, utiliser les fonctions caractéristiques), Si S_n devait converger presque sûrement vers une constante, alors on aurait la loi de S_n qui devrait tendre vers un Dirac en* 0 *(car la loi de Cauchy est symétrique même si $X \notin L^1$), ce qui est faux, voila donc un exemple où la loi forte ne marche pas.*

Cet exemple est dans le livre de Barbe et Ledoux, "Probabilités".

Il existe un très grand nombre de versions des théorèmes limites avec hypothèses affaiblies, la littérature en probabilités en est pleine, certaines peuvent avoir leur place dans cette leçon, néanmoins on se gardera d'aborder (inutilement) des questions trop difficiles (sauf si l'on est sûr de soi), de plus, cela peut rapidement devenir hors-sujet.

La loi forte des grands nombres avec le (très important) lemme de Borel-Cantelli peut fournir un développement, mais cela reste compliqué et ne peut pas être improvisé. Le théorème central-limite ne peut pas être proposé en développement (trop court).

On précisera que comme dans toutes les leçons de probabilités, une bonne compréhension de la définition, de la hiérarchie et de l'explicitation des convergences est indispensable. En effet, dans la loi forte des grands nombres il est essentiel de savoir traduire la convergence presque sûre en terme d'évenements (mais aussi : définir la convergence en probabilités, la convergence en loi, lien avec les normes L^p(utiliser Markov) ...).

Un exercice possible à l'oral est la démonstration de la loi forte des grands nombres quand $E(X^4) < +\infty$. Pour voir cet exercice relativement facile mais néanmoins fondamental (cela intervient dans la démonstration du cas général), on renverra au livre de Barbe et Ledoux, "Probabilités".

On peut rajouter comme application du théorème centrale limite, ou peut-être en partie supplémentaire des éléments de statistiques : détermination d'intervalle de confiance, asymptotique, cela reste dans le cadre de la leçon.

5.4 Variables aléatoires gaussiennes, vecteurs gaussiens

Autre élément indispensable d'une leçon sur l'indépendance, notamment la réciproque non corélation implique indépendance dans le cas de vecteurs gaussiens est incontournable et doit être mise en valeur à l'oral.

Sans rentrer dans les détails, car c'est un sujet très connu et en général très bien traité avec un chapitre particulier (dans le livre de Barbe et Ledoux ou dans le livre d'Ouvrard par exemple), on indique que le théorème de Bernstein (dans le plan), dont la démonstration est à trouver dans le livre d'Ouvrard, fournit un trés bon développement qui est de plus une application non triviale de la fonction caractéristique (voir la leçon sur l'utilisation des transformations de Fourier et de Laplace). On précise également qu'il y aura, à l'oral, une question sur les vecteurs gaussiens ou les variables aléatoires gaussiennes, on ne peut pas se permettre de passer à côté de ce sujet.

Pour l'entrainement on peut considérer le problème suivant (Y_n) échantillon de loi normale centrée de variance 1, et X_0 de loi normale d'espérance α_0 et de variance σ_0^2. Pour tout n $X_{n+1} = aX_n + bY_{n+1}$ où a et b sont des réels. Etudier la convergence de (X_n).

Cet exercice est à voir dans le livre "Modélisation stochastique" de Scheffei-Berka.

5.5 Espérance conditionnelle

Un sujet qui n'a pas trouvé sa place dans notre leçon car jugé un peu trop spécifique et un peu compliqué à introduire. Le problème étant que dans cette leçon sur l'indépendance, si on veut parler d'espérance conditionnelle, on ne peut pas se passer de la définir (on pourrait s'en passer dans des leçons plus portées sur les exemples, variables de Bernoulli...). On est alors devant un problème difficile à résoudre, faut-il en donner une définition simple mais pas assez théorique (voir "Les maths pour l'agreg") ce qui pose des problèmes si l'on veut aborder des problèmes compliqués, ou faut-il alors donner une définition compliquée et prendre le risque de s'égarer dans des détails largement inutiles, au mieux, voire carrément hors-sujet, au pire (sans parler de la possibilité de se tromper, tout simplement) ?

Néanmoins, si l'on est bien au courant de ce problème, et que l'on y a un peu réfléchi, on peut introduire la notion dans cette leçon puisque l'indépendance est très utile quand on considère l'espérance conditionnelle. Cela permet ensuite de considérer les martingales, chaines de Markov, filtrations...

Quelque soit le choix que l'on fait, de parler, ou pas, d'espérance conditionnelle, il faut s'attendre à des questions sur l'espérance conditionnelle, et donc revoir (même succintement) quelques points sur l'espérance conditionnelle.

5.6 Constructions théorique et effective de suites de variables indépendantes

C'est une question qui se pose de manière tout à fait naturel dans cette leçon, il n'est pas nécessaire d'en faire une partie du plan (on peut néanmoins le faire) mais il faut s'être posé la question avant d'aller à l'oral.

Le premier problème est la construction théorique de suites de variables indépendantes d'une même loi (on appelera dans cette partie échantillon d'une loi cette suite). La question est résolue et c'est le théorème d'extension de Kolmogorov qui nous permet d'assurer théoriquement l'existence d'une telle suite pour toute loi. On peut le trouver à la page 96 du livre "Probabilités" de Barbe et Ledoux.

Le second problème s'énonce de la façon suivante, je dispose d'une variable aléatoire de loi uniforme sur $[0,1]$ sur une machine, comment simuler un échantillon fini (nécessairement) d'une loi $\mathcal{L}$ sachant que je connais sa fonction de répartition.

La première chose à voir est que le théorème de Glivencko-Cantelli nous assure qu'il suffit de savoir simuler un nombre fini de variables aléatoires de loi uniforme sur $[0,1]$ indépendantes pour créer une variable aléatoire dont la loi est "arbitrairement proche" de la loi $\mathcal{L}$, cela se fait via la fonction de répartition empirique (important).

Supposons que N lois uniformes nous donnent une précision satisfaisante par ce biais là. Si l'on veut simuler n variables aléatoires indépendantes de loi $\mathcal{L}$ (avec cette précision), il faut donc simuler nN variables aléatoires de loi uniforme sur $[0,1]$.

La question se résume donc à "Comment créer M variables aléatoires de lois uniformes sur $[0,1]$ à partir d'une variable aléatoire de loi uniforme U sur $[0,1]$?"

Définition 12 (i-ème élement du développement dyadique) *Soit $x \in [0,1]$, on sait que*

$x = \sum_{i>0} \frac{1}{2^i} u_i(x)$, *où* $u_i(x) \in \{0,1\}$.

Alors on définit $u_i : [0,1] \to u_i(x)$.

Cet élément va nous ammener à la propriété suivante :

Propriété 17 (Bernoulli dans la loi uniforme) *Pour tout $i > 0$ la variable aléatoire $u_i(U)$ est une loi de Bernoulli de paramètre $\frac{1}{2}$.*

De plus, $(u_i(U))_{i>0}$ est une suite de variables aléatoires indépendantes.

On sait maintenant simuler un nombre arbitrairement grand de variables aléatoires de Bernoulli. Il faut maintenant pouvoir, à partir de cela, simuler un certain nombre de variables aléatoires indépendantes de même loi uniforme sur $[0,1]$. Pour cela on a besoin de la propriété (bien connue) suivante :

Propriété 18 (Equipotence de $\mathbb{N}$ et $\mathbb{N}^2$)

$$\mathbb{N} \leftrightarrow \mathbb{N}^2$$

La bijection associée est

$$f(p,q) = (\sum_{k=0}^{p+q} k) + q$$

La démonstration de cela est faite dans "Les contre-exemples en mathématiques" de Hauchecorne.

Maintenant on construit pour tout $1 \leq j \leq M$, $(X_i^j)_i$,une suite de variables aléatoires indépendantes par :

$$X_i^j = u_{f(j,i)}(U), 1 \leq j \leq M, \forall i$$

On a évidemment $\left\{X_i^j, \, \forall j \forall i\right\}$ est une famille indépendante de variables aléatoires.

Pour terminer on a la dernière propriété :

Propriété 19 (Uniforme par les Bernoulli) *Si (X_i) est un échantillon de Bernoulli de paramètre $\frac{1}{2}$, alors, si :*

$$V = \sum_{i>0} \frac{1}{2^i} X_i$$

V est une variable aléatoire de loi uniforme.

On peut donc construire $(V^j)_{1 \leq j \leq M}$ (par $V^j = \sum_{i>0} \frac{1}{2^i} X_i^j$), un échantillon fini de loi uniforme et notre problème est résolu avec Glivencko-Cantelli.

Une telle discussion ne peut probablement pas faire un développement (quoique...), ni même une partie consistante du plan, néanmoins, ne pas pouvoir justifier l'existence théorique ou la construction effective de suites de variables aléatoires indépendantes identiquement distribuées serait particulièrement malvenu (mais pas catastrophique) dans cette leçon et dans à peu prés toutes les leçons de probabilités.

Cette discussion, pas tout à fait centrale dans cette leçon, prend tout son sens dans la leçon sur les variables aléatoires de Bernoulli.

Le théorème de Glivencko-Cantelli doit être connu (au moins avec les mains) pour toutes les leçons de probabilités.

5.7 Lois infiniment divisibles

Dans notre leçon, nous avons choisi de parler de lois infiniment divisibles. Ce point est le seul point non-essentiel du plan.

Il y a deux éléments qui permettent de faire le lien avec l'indépendance. D'une part, la définition de ces lois infiniment divisibles contient de l'indépendance, d'autre part si un "certain genre" de suite (dans ce genre, il est question d'indépendance) converge en loi vers $\mathcal{L}$ alors $\mathcal{L}$ est une loi infiniment divisible.

Dans la dernière sous-partie de la deuxième partie, on donne, en plus d'une caractérisation via la fonction caractéristique, une formule générale de la fonction caractéristique d'une loi infiniment divisible. Il vaut mieux préciser, si on la donne, qu'elle est ici à titre culturel mais qu'elle n'est pas du tout au niveau de l'agrégation. L'intérêt pédagogique d'une telle formule étant de faire apparaître les deux composantes fondamentales des lois infiniment divisibles : loi normale et loi de Poisson.

Dans la première sous-partie de la troisième partie on évoque le théorème sur la convergence en loi de sommes de variables aléatoires indépendantes et identiquement distribuées. Toutes les lois limites de telles convergence sont alors des lois infiniment divisibles. On remarque alors que toutes les convergences des théorèmes classiques de convergence sont dans ce cadre là (convergence en probabilités et convergence presque-sûre impliquent chacune la convergence en loi). On peut alors y voir là un moyen d'expliquer ou de donner un éclairage théorique ou fondamental sur l'apparition de la loi normale dans le théorème central-limite, alors que le simple développement limité des fonctions caractéristiques (qui est la preuve du théorème centrale-limite) reste tout à fait calculatoire sans dévoiler quoique ce soit.

Une discussion sur les lois infiniment divisibles peut trouver sa place dans une leçon sur les applications de la transformée de Fourier et la transformée de Laplace ou une leçon sur les théorèmes de convergence.

Tout ce qui a été écrit dans le plan sur les lois infiniment divisibles (à part la formule compliquée) peut se trouver dans le livre de Cotterel (et cie). On prendra garde au fait que Cotterel appelle ces lois, des lois inDEfiniment divisibles. Mais, en regardant de près la définition on constate que la division n'est pas inDEfinie mais infinie, il a donc semblé tout à fait logique d'écrire infiniment divisible.