INDEPENDANCE D'EVENEMENTS ET DE VARIABLES ALEATOIRES EXEMPLES.

Cadre : $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilité.

I Définitions et critères d'indépendance. [Barbe]

Def * évènements, tribus : Une famille $(A_i)_{i\in I}$ d'évènements de $\mathcal{F}$ est indépendante si pour tout $J \subset I$, J fini, $\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j)$.

- Une famille $(\mathcal{F}_i)_{i\in I}$ de sous-tribus (ou d'algèbres) de $\mathcal{F}$ est indépendante si toute famille $(A_i)_{i\in I} \subset (\mathcal{F}_i)_{i\in I}$ est indépendante.

* variables aléatoires :

Une famille $(X_i)_{i\in I}$, $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ est indépendante si $(\sigma(X_i))_{i\in I}$ est une famille de sous-tribus de $\mathcal{F}$ indépendante ie : $\forall J$ fini $\subset I$, $\forall (B_i)_{i\in J} \in \mathcal{B}(\mathbb{R}^n)^J$,

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

* Critères d'indépendance :

prop : Soit $(X_1 \cdots X_n)$ une famille finie de v.a.r. Il y a équivalence entre :

(i) $(X_1 \cdots X_n)$ sont des var indépendantes.

(ii) $\mathbb{P}_{(X_1 \cdots X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$

(iii) $\forall (f_i)_{i\in [\![1,n]\!]}$ mesurables bornées $E\left(\prod_{i=1}^n f_i(X_i)\right) = \prod_{i=1}^n E(f_i(X_i))$.

(iv) $\forall (t_1 \cdots t_n) \in \mathbb{R}^n$, $\varphi_{(X_1\cdots X_n)}(t_1 \cdots t_n) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n)$

où $\varphi_X(t) = E(e^{i\langle t, X\rangle})$: fonction caractéristique de X.

(v) $\forall (t_1 \cdots t_n) \in]-1,1[^n$, $G_{(X_1\cdots X_n)}(t_1 \cdots t_n) = G_{X_1}(t_1) \cdots G_{X_n}(t_n)$

où $G_X(t) = E(t_1^{X_1} \cdots t_n^{X_n})$: fonction génératrice.

* **Exemples** : - tirage avec remise

- Simulation : **prop** [Box-Muller] [Ouv p 67]

Si (U,V) indépendantes de loi $\mathcal{U}([0,1])$ alors

$X = \sqrt{-2\log(U)}\cos(2\pi V)$ et $Y = \sqrt{-2\log(U)}\sin(2\pi V)$

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

II Nombre fini de variables indépendantes. Echantillon.

En statistique on dispose souvent d'un nombre fini de valeurs à étudier. L'idée est de voir ces données comme des réalisations de variables indépendantes et identiquement distribuées.

On appelle échantillon une famille finie de va iid $(X_1 \cdots X_n)$.

L'existence d'un nuplet $(X_1 \cdots X_n)$ de va indépendantes tq $X_i \sim \mu_i$ est facile : $\Omega = \mathbb{R}^n$, muni de $\bigotimes_{i=1}^n \mathcal{B}(\mathbb{R})$ et $\mathbb{P} = \mu_1 \otimes \cdots \otimes \mu_n$.

Alors : les projections $X_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $(x_i) \longmapsto x_i$ sont des var indépendantes de loi μ_i.

* Somme finie de va indépendantes [Barbe]

Def : Si $X \perp\!\!\!\perp Y$ (X et Y indépendants), on appelle convolée de $\mathbb{P}_X$ et $\mathbb{P}_Y$ la loi de $X+Y$. On note $\mathbb{P}_X * \mathbb{P}_Y$.

Prop : pour tout f mesurable bornée, $\int_{\mathbb{R}} f \, d(\mathbb{P}_X * \mathbb{P}_Y) = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(x+y)\mathbb{P}_Y(dy)\right)\mathbb{P}_X(dx)$

Rq : si $\mathbb{P}_X = f\,d\lambda$, $\mathbb{P}_Y = g\,d\lambda$, $\mathbb{P}_X * \mathbb{P}_Y = f*g\,d\lambda$.

Exemples : - $(X_1 \cdots X_n)$ iid, $X \sim \mathcal{B}(p) \Rightarrow X_1 + \cdots + X_n \sim \mathcal{B}(n,p)$.

- $X \sim \mathcal{N}(m_1, \sigma_1^2)$, $Y \sim \mathcal{N}(m_2, \sigma_2^2)$, $X \perp\!\!\!\perp Y \Rightarrow X+Y \sim \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$

- $X \sim \gamma(a_1, \lambda)$, $Y \sim \gamma(a_2, \lambda)$, $X \perp\!\!\!\perp Y \Rightarrow X+Y \sim \gamma(a_1+a_2, \lambda)$.

Prop : Si X, Y, var, $X \perp\!\!\!\perp Y \Rightarrow E(XY) = E(X)E(Y)$

et donc $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Ex : Si $X \sim \mathcal{B}(n,p)$ $E(X) = np$ et $\mathrm{Var}(X) = np(1-p)$.

Cor: $(X_1 \dots X_n)$ iid, $E(X^2) < +\infty$, $S_n = X_1 + \dots + X_n$, $\varepsilon > 0$. Alors :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - E(X)\right|^2 \geq \varepsilon\right) \leq \frac{Var(X)}{n\varepsilon^2}.$$

[Dev] p50 * **Statistique d'ordre** : $(X_1, \dots X_n)$ n-échantillon de loi μ telle que $\mathbb{P}(X_i = X_j) = 0$ pour tout $i \neq j$. Le n-uplet réordonné $(X_{\sigma(1)} < \dots < X_{\sigma(n)})$ est appelé statistique d'ordre de $(X_1 \dots X_n)$. La loi de $Y = (X_{\sigma(1)} < \dots < X_{\sigma(n)})$ est égale à $n!\, \mathbb{1}_{(x_1 < \dots < x_n)}\, \mu^{\otimes n}$. σ est une va uniforme sur S_n et Y et σ sont indépendantes.

Rq: On note $(X_{(1)} \dots X_{(n)}) = (X_{\sigma(1)} \dots X_{\sigma(n)})$.

$X_{(1)} = \min(X_1 \dots X_n)$, $X_{(n)} = \max(X_1 \dots X_n)$.

[FFI p189] **Exemple:** * Si $(X_1 \dots X_n) \sim (\mathcal{U}_{[0,1]})^{\otimes n}$ alors $X_{(1)} \sim Beta(1, n)$, $X_{(n)} \sim Beta(n, 1)$ où $Beta(r, s) = \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1} \mathbb{1}_{[0,1]}(x)\,dx$.

* $X \sim \mathcal{E}(\lambda)$, $Y \sim \mathcal{E}(\beta)$, $X \perp Y \Rightarrow \min(X, Y) \sim \mathcal{E}(\lambda + \beta)$.

[ouv] * **Vecteurs gaussiens.**

Rq: $X \perp Y \Rightarrow E(XY) = E(X)E(Y)$ mais la réciproque est fausse

cf: $X \sim \mathcal{N}(0,1)$, $Y = X^2$,

Néanmoins: **prop** : Soit $X = (X_1 \dots X_n)$ vecteur gaussien. Alors les X_i sont gaussiennes et elles sont indépendantes ssi elles sont 2 à 2 non corrélées.

Rq: c'est faux si on ne suppose rien sur X : **cf:** $X_1 \sim \mathcal{N}(0,1)$, $X_2 = \varepsilon X_1$ où $\varepsilon \sim \mathcal{R}(1/2)$, $\varepsilon \perp X_1$.

[Cot p211] **Prop:** Soient $(X_1 \dots X_n)$ iid de carré intégrable, $m = E(X_1)$, $\sigma^2 = Var(X_1)$, $\overline{X} = \frac{1}{n}(X_1 + \dots + X_n)$, $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left((X_i - \overline{X})^2\right)$.

Alors $X_1 \sim \mathcal{N}(m, \sigma^2) \Leftrightarrow \overline{X} \perp S^2$.

Application : test de **Student** pour l'estimation de la moyenne d'une gaussienne de variance inconnue. [ouv] p 295

Th: (Indécomposabilité de Cramer).

$$\left.\begin{cases} X \perp Y \\ X + Y \text{ gaussienne} \end{cases}\right\} \Longrightarrow X \text{ gaussienne et } Y \text{ gaussienne.}$$

[ZQ] p 503

III Suite de variables aléatoires.

L'existence d'une suite de var indépendantes est moins claire que dans le cas fini.

Th: [Construction de va iid $\sim B(1/2)$] [ouv p 53]

Pour tout $\omega \in [0,1[$, soit $\omega = \sum_{n=1}^{+\infty} \frac{D_n(\omega)}{2^n}$ son développement dyadique propre. Alors sur $([0,1], \mathcal{B}([0,1]), \lambda_{|[0,1]})$, D_n est une suite de va iid de loi $B(1/2)$.

Cor: Soit $(\mu_i)_{i \in \mathbb{N}}$ une suite de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il existe une suite de var (X_j) définies sur $([0,1], \mathcal{B}([0,1]), \lambda_{|[0,1]})$ indépendantes et pour tout $j \in \mathbb{N}$, $X_j \sim \mu_j$.

* **Exemples de suites de va.**

- **loi du premier succès au pile ou face** : Soit (X_n) iid $\sim B(p)$. Alors $T = \inf\{n \in \mathbb{N}, X_n = 1\}$ suit une loi $\mathcal{G}(p)$.

- **Processus de Poisson** : [ouv p171]

Soit (W_n) une suite iid de loi $\mathcal{E}(\lambda)$. $(\lambda > 0)$.

$T_n := \sum_{k=1}^{n} W_k$ et $N_t := \sum_{n \in \mathbb{N}^*} \mathbb{1}_{(T_n \leq t)}$ pour $t \geq 0$.

$(N_t)_{t \in \mathbb{R}^+}$ est appelé **processus de Poisson.**

Prop * $N_t \sim \mathcal{P}(\lambda t)$

* Conditionnellement à $\{N_t = n\}$, $(T_1, \dots, T_n) \sim \frac{n!}{t^n} \mathbb{1}_{\{0<t_1<\dots<t_n\}} d(t_1 \dots t_n)$

↳ on retrouve la loi de la statistique d'ordre d'un échantillon de loi $\mathcal{U}([0,1])$ (loi de Dirichlet).

a) Théorème de Borel-Cantelli. [Bar]

Déf: (A_n) une suite d'événements : on définit

$$\limsup (A_n) = \bigcap_{n\in\mathbb{N}} \bigcup_{p\geq n} A_p \text{ et } \liminf A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{p\geq n} A_p.$$

Th: [Borel-Cantelli] Soit (A_n) une suite d'événements:

Alors: (i) si $\sum P(A_n) < +\infty$ alors $P(\limsup A_n) = 0$.

(ii) si $\sum P(A_n) = +\infty$ alors $P(\limsup A_n) = 1$, si les (A_n) sont indépendants.

Rq: (i) n'a rien à voir avec l'indépendance.

(ii) est fausse sans l'indépendance : cf $A_n =]0, \frac{1}{n}]$ dans $\Omega = [0,1]$.

[Bar] Applications * On joue à pile ou face une infinité de fois. Alors "deux piles consécutifs" apparaissent une infinité de fois de façon presque sûre.

* loi faible des grands nombres pour $E(X^4) < +\infty$.

[DMK] **Prop** [Marche aléatoire sur $\mathbb{Z}^d$]

Soit (S_n) la marche aléatoire sur $\mathbb{Z}^d$. Alors :

* si $d \geq 3$ $P(\limsup (S_n = 0)) = 0$: transience

* si $d = 1$ ou 2 $P(\bigcup_{n\geq 1} (S_n = 0)) = 1$: récurrence.

Rq: la récurrence ne provient pas de Borel-Cantelli.

[Bar] b) Loi du 0-1 ; Applications.

Soit (X_n) suite de variables indépendantes et $\mathcal{F}_n = \sigma(X_k, k\leq n)$.

Un événement est dit asymptotique s'il appartient à la tribu :

$$\mathcal{F}_\infty = \bigcap_{n\in\mathbb{N}} \sigma(\mathcal{F}_p, p \geq n).$$

Prop Si A est asymptotique, $P(A) = 0$ ou 1.

Cor: toute va asymptotique est P ps constante.

* Exemples : - $\limsup A_n$, $\liminf A_n \in \mathcal{F}_\infty$, si $A_n \in \mathcal{F}_n$.

- $\{X_n$ converge$\}$, $\{\sum X_n$ converge$\} \in \mathcal{F}_\infty$.

- $\liminf \frac{1}{n}\sum X_n$ et $\limsup \frac{1}{n}\sum X_n$ sont P ps constantes.

- le rayon de convergence de $\sum X_n z^n$ est P ps constante.

- Th de Steinhaus : si $\sum a_n z^n$ de RCV R, $\exists (\varepsilon_n)$ tq $|a_n| = |\varepsilon_n|$, tq $D(0,R)$ est une coupure pour $\sum \varepsilon_n z^n$. [ZQ] p541.

c) Temps d'arrêt.

Déf: Soit (X_n) suite de va. $\mathcal{F}_n = \sigma(X_k, k\leq n)$. Un temps d'arrêt T est une va à valeurs dans $\mathbb{N}\cup\{+\infty\}$ tq $\forall n\in\mathbb{N}$, $\{T = n\} \in \mathcal{F}_n$.

Th: (Wald) Si (X_n) iid, $E(|X|) < +\infty$, T temps d'arrêt intégrable alors

$$E\left(\sum_{i=1}^{T} X_i\right) = E(T)\,E(X_1). \quad \text{[FF2 p23]}.$$

Applications * ruine du joueur : $X_n \sim R(1/2)$: gain du joueur à la n-ième partie ; a : fortune initiale, b : fortune de la banque ; $S_n = \sum_{i=1}^{n} X_i$: gain accumulé

$T = \inf\{n / S_n = -a$ ou $S_n = b\}$. Alors $P(T = -a) = \frac{b}{b+a}$, $E(S_T) = ab$. [Dac p27]

* méthode de rejet en simulation.

d) Théorèmes asymptotiques.

Th: (loi des grands nombres) : (X_n) iid intégrables, alors $\frac{X_1 + \dots + X_n}{n} \xrightarrow{ps} E(X_1)$

Th: (théorème central limite) : (X_n) iid de carré intégrable : $\sqrt{n}\left(\frac{S_n}{n} - E(X)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

* Application * méthode de Monte Carlo, intervalles de confiance.

prop : presque tout réel de $[0,1]$ est normal [ZQ p. 550]

e) Séries de variables indépendantes

Th: (X_n) indépendantes centrées. Alors $(\sum E(X_n^2) < +\infty \Rightarrow \sum X_n$ converge ps$)$ et la réciproque est vraie si les (X_n) sont uniformément bornées. [ZQ p524]

Exemple $\sum \frac{\varepsilon_n}{n^\alpha}$ converge ps $\Leftrightarrow \alpha > \frac{1}{2}$. ($\varepsilon_n$ iid $\sim R(1/2)$)

Th: (X_n) indépendantes, $S_n = X_1 + \dots + X_n$. [Bil p288]

alors (S_n) converge en probabilité $\Leftrightarrow (S_n)$ converge ps.

Références : [Bar] : Barbe - Ledoux.

[Ouv] : Ouvrard 2.

[FF1] : Fuchs Foata calcul des probabilités

[FF2] : ——— processus stochastiques.

[Cot] : Cottrell ... exercices de probabilités

[ZQ] : A votre avis ...

[Bil] : Billingsley , Probability and measure.

[Dac] : Dacunha castelle Duflo , Exercices problèmes à temps fixe 1.

[Rev] : Revuz (pour la statistique d'ordre).

[Poe] : Park . Surimi and surimi sea-food (pour les temps d'arrêt).

Développements possibles : * Kolmogorov via les dyadiques.

* $S^2 \perp \overline{X}$

* convergence de (S_n)

* marche aléatoire.

* ruine du joueur

* Indécomposabilité de Cramer