260

ESPÉRANCE, VARIANCE ET MOMENTS D'UNE VARIABLE ALÉATOIRE

Cadre : $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé sur lequel X et Y sont des va réelles.

I DÉFINITIONS & PREMIÈRES PROPRIÉTÉS

1 Espérance d'une variable aléatoire

Déf 1 : ([BL], III.4.1)
Si X est intégrable (ie $\int_\Omega |X(\omega)|\, d\mathbb{P}(\omega) < \infty$)
On appelle espérance de X le réel $\mathbb{E}[X] := \int_\Omega X(\omega)\, d\mathbb{P}(\omega) =: \int_\Omega X\, d\mathbb{P}$.

Rq 2 : • Si $\mathbb{E}[X] = 0$, on dit que X est centrée.
• L'espérance d'une va est sa valeur moyenne ; elle est linéaire.

Ex 3 : Si X est constante presque sûrement, alors $\mathbb{E}[X] = X$ ps.

Thm 4 : Théorème de transfert ([BL], III.4.2)
Soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction borélienne.
• Si ϕ est à valeurs positives,
Alors $\mathbb{E}[\phi(X)] = \int_\Omega \phi(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} \phi(x)\, d\mathbb{P}_X(x)$
(où $\mathbb{P}_X$ est la probabilité définie par : $\forall A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_X(A) := \mathbb{P}(X \in A)$.)
• Si ϕ est à valeurs quelconques,
Alors $\phi(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \Leftrightarrow \phi \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$
Et dans ce cas $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x)\, d\mathbb{P}_X(x)$.

Ex 5 : Pour tout $A \in \mathcal{B}(\mathbb{R})$, on a : $\mathbb{E}[\mathbb{1}_A(X)] = \mathbb{P}(X \in A)$. ([BL], III.4.3)

Thm 6 : Inégalité de Jensen ([BL], III.4.5).
Soit $\phi : \mathbb{R} \to \mathbb{R}$ convexe ; on suppose X et $\phi(X)$ intégrables.
Alors $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$.

Thm 7 : Inégalité de Markov ([BL], III.4.9)
Soit $t > 0$; on suppose X intégrable.
Alors $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[|X|]}{t}$.

Déf 8 :
Si $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ est un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ tel que $\forall i \in [\![1,d]\!]$, X_i est intégrable,
Alors on définit son espérance comme $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]) \in \mathbb{R}^d$.

Thm 9 : ([BL], IV.1.11)
Une famille de va réelles $(X_i)_{i \in I}$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ est mutuellement indépendante ssi pour tout $J \subset I$ fini, et pour toute famille de fonctions boréliennes $(\phi_i)_{i \in J}$ telles que $\forall i \in J$, $\phi_i(X_i) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, on a :
$$\mathbb{E}\Big[\prod_{i \in J} \phi_i(X_i)\Big] = \prod_{i \in J} \mathbb{E}[\phi_i(X_i)].$$

Cor 10 :
Si $(X_1, \dots, X_n)$ est mutuellement indépendante, alors $\mathbb{E}[X_1 \dots X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n]$

2 Espérances des lois usuelles

Prop 11 : ([Ouv1], 5.1.1)
Si X est discrète et intégrable,
Alors $\mathbb{E}[X] = \sum_{x \in val(X)} x\, \mathbb{P}(X = x)$ où $val(X) = \{x \in \mathbb{R} \mid \mathbb{P}(X = x) > 0\}$.

Ex 12 : ([Ouv1], Appendice)
• Loi de Bernoulli : si $X \sim b(p)$, alors $\mathbb{E}[X] = p$.
• Loi binomiale : si $X \sim \mathcal{B}(n,p)$, alors $\mathbb{E}[X] = np$
• Loi de Poisson : si $X \sim \mathcal{P}(\lambda)$, alors $\mathbb{E}[X] = \lambda$.
• Loi géométrique : si $X \sim \mathcal{G}(p)$, alors $\mathbb{E}[X] = \frac{1}{p}$.
• Loi uniforme : si $X \sim \mathcal{U}(\{x_1, \dots, x_n\})$, alors $\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i$.

Prop 13 : ([Ouv1], déf 6.8)
Si X admet une densité f_X et si $x \mapsto |x| f_X(x) \in L^1(\mathbb{R})$,
Alors $\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\, dx$.

Ex 14 : ([Ouv1], Appendice)
• Loi uniforme : si $X \sim \mathcal{U}([a,b])$, alors $\mathbb{E}[X] = \frac{a+b}{2}$.
• Loi normale : si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\mathbb{E}[X] = m$.
• Loi exponentielle : si $X \sim \mathcal{E}(\lambda)$, alors $\mathbb{E}[X] = \frac{1}{\lambda}$

C-ex 15 : La loi de Cauchy de densité $f_X : x \mapsto \frac{1}{\pi(1+x^2)}$ n'admet pas d'espérance.

3 Espérance conditionnelle

Déf 16 : ([BL], VI.2.1)
Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$; on suppose X intégrable.
Alors, il existe une p.s. unique va, appelée espérance conditionnelle de X sachant $\mathcal{B}$ et notée $\mathbb{E}[X|\mathcal{B}]$ telle que :
• $\omega \mapsto \mathbb{E}[X|\mathcal{B}](\omega)$ est $\mathcal{B}$-mesurable ;
• $\forall B \in \mathcal{B}$, $\int_B \mathbb{E}[X|\mathcal{B}]\, d\mathbb{P} = \int_B X\, d\mathbb{P}$.

Prop 17 : ([BL], VI.2.2)
L'espérance conditionnelle sachant $\mathcal{B}$ est linéaire et positive.

Prop 18 : ([BL], VI.2.2)
Si $\mathcal{B} \perp\!\!\!\perp \sigma(X)$, alors $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$ ps.

Prop 19 : ([BL], VI.2.2)
Si Y est $\mathcal{B}$-mesurable et XY intégrable, alors $\mathbb{E}[XY|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]\, Y$ ps.

Ex 20 : ([Cot], exo 6.1)
• Si $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$ avec $X_1 \perp\!\!\!\perp X_2$, alors $\mathbb{E}[X_1 | X_1 + X_2] = \frac{n_1(X_1+X_2)}{n_1+n_2}$ ps
• Si $X_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 \sim \mathcal{P}(\lambda_2)$ avec $X_1 \perp\!\!\!\perp X_2$, alors $\mathbb{E}[X_1 | X_1 + X_2] = \frac{\lambda_1(X_1+X_2)}{\lambda_1+\lambda_2}$ ps.

II MOMENTS D'UNE VARIABLE ALÉATOIRE

Déf 21: ([BL], III.4.1)
Si $\int_\Omega |X|^p \, dP < \infty$, alors on définit le moment d'ordre p de X par :
$$\mathbb{E}[X^p] = \int_\Omega X^p \, dP$$

1 Moment d'ordre 2 : variance & covariance

Déf 22: ([BL], III.4.6)
Si X est de carré intégrable, on appelle variance de X le réel :
$var(X) = \mathbb{E}[(X-\mathbb{E}[X])^2]$; l'écart-type de X est $\sigma(X) = \sqrt{var(X)}$.

Rq 23: • On peut écrire $var(X) = \|X-\mathbb{E}[X]\|^2_{L^2(\Omega,\mathcal{A},P)}$. ([BL], III.4.6)
• On a l'égalité $var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Ex 24: • Si $var(X) = 0$, alors $X = \mathbb{E}[X]$ ps.
• Si $X \sim \mathcal{B}(n,p)$, alors $var(X) = np(1-p)$. ([BL], III.4.7)
• Si $X \sim \mathcal{N}(0,1)$, alors $var(X) = 1$

Prop 25: ([BL], III.4.7)
Si $\alpha \in \mathbb{R}$, on a : $var(X+\alpha) = var(X)$ et $var(\alpha X) = \alpha^2 var(X)$.

Thm 26: Inégalité de Tchebychev ([BL], III.4.10)
Si $X \in L^2$, alors $P(|X-\mathbb{E}[X]| \geq t) \leq \frac{var(X)}{t^2}$ pour tout $t > 0$.

Thm 27: Inégalité de Cauchy-Schwarz ([BL], III.4.5)
Si $X \in L^2$ et $Y \in L^2$, alors $XY \in L^1$ et $\mathbb{E}[|XY|]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$.

Déf 28: ([BL], III.4.11)
Si $X, Y \in L^2$, on définit leur covariance $cov(X,Y) := \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$

Ex 29: ([BL], III.4.12)
Si (X,Y) admet une densité $f_{X,Y} : (x,y) \mapsto \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)$, alors $cov(X,Y) = 0$

2 Lien avec l'indépendance

Déf 30: ([BL], IV.1.13)
Si $X, Y \in L^2(\Omega,\mathcal{A},P)$, on dit que X et Y sont non-corrélées ssi
$cov(X,Y) = 0$ (ce qui équivaut à $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).

Rq 31: L'indépendance implique la non-corrélation. ([BL], IV.1.14)

C-ex 32: Soit $X \sim \mathcal{N}(0,1)$ et $Y = X^2$.
Alors X et Y sont non-corrélées, mais pas indépendantes. ([BL], IV.1.14)

Thm 33:
Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien dans $\mathbb{R}^d$.
Si les composantes de X sont deux à deux non-corrélées,
Alors la famille $(X_1, \ldots, X_d)$ est mutuellement indépendante.

Prop 34: Identité de Bienaymé ([BL], IV.1.15)
Si $X_1, \ldots, X_n$ sont de carré intégrable et 2 à 2 non-corrélées,
Alors $var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n var(X_i)$.

App 35: Si $X \sim \mathcal{B}(n,p)$, alors $var(X) = np(1-p)$.

App 36: Théorème de Weierstrass ([ZQ], XIII.II.1.c)
Soit $f : [0,1] \to \mathbb{C}$ continue et $\omega : h \mapsto \sup\{|f(u)-f(v)| \mid |u-v| < h\}$.
Pour $n \in \mathbb{N}^*$, on pose $B_n = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$.
Alors :
1) (B_n) converge uniformément vers f sur [0,1] et $\exists C > 0, \forall n \in \mathbb{N}^*, \|f-B_n\|_\infty \leq C\omega\left(\frac{1}{\sqrt{n}}\right)$
2) Cette majoration est optimale, au sens où pour une certaine fonction f lipschitzienne, $\exists \delta > 0, \forall n \in \mathbb{N}^*, \|f-B_n\|_\infty \geq \delta\omega\left(\frac{1}{\sqrt{n}}\right)$.

3 Moments d'ordre p.

Thm 37: Inégalité de Hölder ([BL], III.4.5)
Si $X \in L^p$ et $Y \in L^q$, avec $p, q \geq 1$ et $\frac{1}{p} + \frac{1}{q} = 1$,
Alors $XY \in L^1$ et $\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \, \mathbb{E}[|Y|^q]^{1/q}$.

Prop 38: ([BL], III.4.5)
L'application $r \mapsto \mathbb{E}[|X|^r]^{1/r}$ est croissante.

App 39: ([Cot], 5.21)
Si $1 \leq p \leq q$, alors $L^q(\Omega,\mathcal{A},P) \subset L^p(\Omega,\mathcal{A},P)$ et la convergence dans L^q implique la convergence dans L^p.

Prop 40: ([BL], III.4.8)
On suppose ici que X est une va positive.
Alors, pour tout $p \in]0,+\infty[$, on a : $\mathbb{E}[X^p] = p \int_0^{+\infty} t^{p-1} P(X > t) \, dt$.

App 41: ([BL], III.4.8)
On en déduit un critère d'intégrabilité :
$\mathbb{E}[|X|] < \infty \Leftrightarrow$ pour un ou tout $\varepsilon > 0$, $\sum_{n=0}^\infty P(|X| > \varepsilon n) < \infty$ ou $\sum_{n=0}^\infty 2^n P(X > \varepsilon 2^n) < \infty$.

III UTILISATION DES MOMENTS

1 Fonction génératrice

Dans cette sous-partie, X est à valeurs dans $\mathbb{N}$.

Déf 42: ([Ouv1], 5.23)
Pour $s \in \mathbb{R}$, on pose $G_X(s) = \mathbb{E}[s^X]$ quand $s^X \in L^1(\Omega,\mathcal{A},P)$.
G_X est appelée fonction génératrice de X ; elle est bien définie sur [-1,1].

Prop 43: ([Ouv1], 5.24)
1) $\forall s \in [-1,1]$, $|G_X(s)| \leq 1$ et $G_X(1) = 1$.
2) $\forall s \in [-1,1]$, $G_X(s) = \sum_{n=0}^\infty P(X=n) s^n$.
3) G_X est continue sur [-1,1] et C^∞ sur]-1,1[.

Thm 44: ([Ouv1], 5.24)
G_X caractérise la loi de X; en effet : $\mathbb{P}(X=n) = \frac{G_X^{(n)}(0)}{n!}$.

Ex 45: • Si $X \sim \mathcal{B}(n,p)$, alors $G_X(s) = (ps+1-p)^n$, $s\in[-1,1]$ ([Ouv1], 5.3.2)
• Si $X \sim \mathcal{P}(\lambda)$, alors $G_X(s) = \exp(\lambda(s-1))$, $s\in[-1,1]$

Prop 46: ([Ouv1], 5.25)
Soient X et Y deux va indépendantes à valeurs dans $\mathbb{N}$.
On a : $\forall s\in[-1,1]$, $G_{X+Y}(s) = G_X(s)G_Y(s)$.

App 47: ([Cot], exo 3.2)
On ne peut pas truquer deux dés indépendants de façon que la somme des points obtenue en les lançant soit équirépartie.

Prop 48: ([Ouv1], 5.27)
X admet un moment d'ordre $r\in\mathbb{N}^*$
$\Leftrightarrow G_X$ est r fois dérivable à gauche en 1.
Dans ce cas, $G_X^{(r)}(1^-) = \sum_{k=r}^{\infty} k(k-1)\dots(k-r+1)\,\mathbb{P}(X=k) = \mathbb{E}[X(X-1)\dots(X-r+1)]$
En particulier, $\mathbb{E}[X] = G_X'(1^-)$.

DÉVELOPPEMENT N°1.

Thm 49: Processus de Galton-Watson.
Soit X une va intégrable à valeurs dans $\mathbb{N}$.
On note $p_k = \mathbb{P}(X=k)$, $m = \mathbb{E}[X]$ et on construit (Z_n) par $\begin{cases} Z_0 = 1 \\ Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n} \end{cases}$,
où les $X_{i,j}$ sont des vaiid de loi $\mathbb{P}_X$.
Alors :
1) G_X est C^1 sur $[0,1]$, strictement croissante sur $]0,1[$, convexe sur $]0,1[$ et même strictement convexe sur $]0,1[$ ssi $p_0+p_1<1$.
2) $\forall n\in\mathbb{N}^*$, $G_{Z_n} = \underbrace{G\circ\dots\circ G}_{n \text{ fois}}$ sur $[0,1]$.
3) la probabilité d'extinction $\pi_\infty := \mathbb{P}(\exists n\in\mathbb{N}, Z_n=0)$ est le plus petit point fixe de G_X sur $[0,1]$.
4) Si $m\leq 1$, alors $\pi_\infty = 1$; sinon π_∞ est l'unique point fixe de G_X sur $]0,1[$.

2 Fonction caractéristique & Transformée de Laplace

Déf 50: ([BL], III.5.1)
On appelle fonction caractéristique de X la fonction $\varphi_X : \begin{vmatrix} \mathbb{R} \to \mathbb{C} \\ t \mapsto \mathbb{E}[e^{itX}] \end{vmatrix}$.

Thm 51: ([BL], III.5.2)
Si X et Y sont deux va telles que $\varphi_X = \varphi_Y$, alors $\mathbb{P}_X = \mathbb{P}_Y$.

Ex 52: • Si $X \sim \mathcal{N}(0,1)$, alors $\varphi_X(t) = \exp\left(-\frac{t^2}{2}\right)$ ([BL], III.5.3)
• Si $X \sim \mathcal{P}(\lambda)$, alors $\varphi_X(t) = \exp(\lambda(e^{it}-1))$

Thm 53: ([BL], III.5.6)
• Si $\mathbb{E}[|X|^n]<\infty$, alors φ_X est n fois dérivable et $\varphi_X^{(n)}(t) = i^k\,\mathbb{E}[X^k e^{itX}]$.
• Si n est pair et si φ_X est n fois dérivable en 0,
Alors X admet un moment d'ordre n.

Prop 54: ([Ouv2], 12.16)
Si X admet des moments de tout ordre et si $\limsup_{n\to\infty} \frac{\mathbb{E}[|X|^n]^{1/n}}{n} < \infty$,
Alors φ_X est analytique.

Thm 55: ([BL], III.5.7)
Si φ_X est analytique, alors $\mathbb{P}_X$ est caractérisée par $(\mathbb{E}[X^k])_{k\in\mathbb{N}}$.

C-ex 56: ([BL], exo III.7)
Si $X\sim\mathcal{N}(0,1)$, la densité de $Z := e^X$ est $f_Z(z) = \frac{1}{\sqrt{2\pi}\,z}\exp\left(-\frac{(\ln z)^2}{2}\right)\mathbb{1}_{\mathbb{R}_+^*}(z)$.
Pour $a\in[-1,1]$, soit $f_a(z) = f_Z(z)(1+a.\sin(2\pi \ln z))$.
Si Z_a est de densité f_a, alors Z_a et Z ont mêmes moments.

Déf 57: ([BL], III.5.9)
On appelle transformée de Laplace de X la fonction $L_X : t\mapsto \mathbb{E}[e^{tX}]$, définie quand $e^{tX}\in L^1$.

App 58: Inégalité de Hoeffding ([Ouv2], exo 10.11)
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de va réelles, indépendantes et centrées.
On suppose que $\forall n\in\mathbb{N}^*$, X_n est ps bornée par une constante c_n.
On note $S_n = \sum_{j=1}^n X_j$ et $a_n = \sum_{j=1}^n c_j^2$.
Alors :
$\forall\varepsilon>0, \forall n\in\mathbb{N}^*$, $\mathbb{P}(|S_n|>\varepsilon) \leq 2\exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.

Cor 59:
Soit $\alpha>0$; on ajoute l'hypothèse : $\exists\beta>0$, $\forall n\in\mathbb{N}^*$, $a_n \leq n^{2\alpha-\beta}$
Alors $\frac{S_n}{n^\alpha} \xrightarrow[n\to\infty]{ps} 0$.

DÉVELOPPEMENT N°2

App 60: Obtention d'un intervalle de confiance par excès ([CV], 3.2)

Thm 61: ([BL], III.5.9)
Si X et Y sont deux va telles que $L_X = L_Y$ dans un voisinage de 0, alors $\mathbb{P}_X = \mathbb{P}_Y$.

Thm 62:
Si $e^{tX}\in L^1$ pour tout t dans un intervalle ouvert contenant 0,
Alors L_X est analytique dans un voisinage V de 0 et $L_X(t) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}[X^n]$, $t\in V$.

3 Convergence

Thm 63: Loi faible des grands nombres ([BL], V.5.1)
Si $\mathbb{E}[|X|]<\infty$, alors $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X]$ où $(X_i)_{i\geq 1}$ iid de loi $\mathbb{P}_X$.

Rq 64: LGN forte (admise) : $\mathbb{E}[|X|]<\infty \Leftrightarrow \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{ps} \mathbb{E}[X]$ ([BL], V.5.2)

Rq 65: Cela signifie que la moyenne empirique est un estimateur fortement consistant de l'espérance.

Thm 66: Théorème central limite. ([BL], V.5.4)
Si $\mathbb{E}[X^2]<\infty$, et si $var(X) =: \sigma^2 > 0$,
Alors $\frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^n X_i - n\mathbb{E}[X]\right) \xrightarrow[n\to\infty]{\mathcal{L}} Z$, où $Z\sim\mathcal{N}(0,1)$.

([BL], V.5.5)

App 67: Intervalle de confiance asymptotique pour le paramètre d'une loi de Bernoulli

Références :

[BL] : P. Barbe & M. Ledoux, Probabilité

[Ouv1] : J.-Y. Ouvrard, Probabilités 1

[Ouv2] : J.-Y. Ouvrard, Probabilités 2

[Cot] : M. Cottrell, V. Genon-Catalot, C. Duhamel & T. Meyre, Exercices de probabilités.

[ZQ] : H. Queffélec & C. Zuily, Analyse pour l'agrégation.

[CV] : B. Cadre & C. Vial, Statistique mathématique.

Processus de Galton-Watson

Pas de véritable référence bibliographique.

Soit X une variable aléatoire intégrable à valeurs dans $\mathbb{N}$.
On note, pour $n \in \mathbb{N}$, $p_n = \mathbb{P}(X = n)$ et $m = \mathbb{E}[X] = \sum_{n=0}^{\infty} np_n < \infty$.
Soit $(X_{i,j})_{i,j\in\mathbb{N}}$ une famille de variables aléatoires indépendantes et identiquement distribuées, suivant la loi $\mathbb{P}_X$.
On définit la suite $(Z_n)_{n\in\mathbb{N}}$ de la façon suivante :

$$\begin{cases} Z_0 = 1 \\ \forall n \in \mathbb{N}, Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n} \end{cases}$$

L'idée est alors de modéliser avec (Z_n) la taille d'une population ; plus précisément, Z_n symbolisera le nombre d'individus à la $n^{\text{ème}}$ génération, et pour $i \in [\![1, Z_n]\!]$, $X_{i,n}$ représentera le nombre de descendants que l'individu de la $n^{\text{ème}}$ génération portant le numéro i aura engendré (les individus de la population qu'on considère génèrent des enfants tous seuls).
On va étudier la suite (Z_n), et particulièrement, on va répondre à la question "Que vaut $\mathbb{P}(\exists n \in \mathbb{N}, Z_n = 0)$? (Quelle est la probabilité que la population considérée s'éteigne ?)" [1]

Lemme

On a : $\forall n \in \mathbb{N}^*, \forall i \in \mathbb{N}, Z_n \perp\!\!\!\perp X_{i,n}$.

Démonstration :

Soit $n \in \mathbb{N}^*$; Z_n ne dépend que de Z_{n-1} et de la famille $(X_{i,n-1})_{i\in\mathbb{N}}$.
Ainsi, par une récurrence immédiate, il vient : Z_n ne dépend que de la famille $(X_{i,j})_{i\geqslant 0, j<n}$.
Et, par indépendance des variables $X_{i,j}$, on obtient que $\forall i \in \mathbb{N}, Z_n \perp\!\!\!\perp X_{i,n}$. ■

On note, pour $n \in \mathbb{N}$, $\pi_n = \mathbb{P}(Z_n = 0)$; et $\pi_\infty = \mathbb{P}(\exists n \in \mathbb{N}, Z_n = 0)$ la probabilité d'extinction.
Comme $Z_n = 0 \Rightarrow Z_{n+1} = 0$, la suite d'événements $(\{Z_n = 0\})_{n\in\mathbb{N}}$ est croissante et on a bien :

$$\pi_\infty = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \{Z_n = 0\}\right) = \lim_{n\to\infty} \pi_n$$

Si $p_0 = 0$, alors on a $\forall n \in \mathbb{N}^*, Z_n \geqslant 1$ ps et $\pi_\infty = 0$.
Si $p_0 = 1$, alors on a $\forall n \in \mathbb{N}^*, Z_n = 0$ ps et $\pi_\infty = 1$.
On suppose donc désormais $p_0 \in]0, 1[$.

1. On aurait pu se poser la question "Que vaut $\mathbb{E}[Z_n]$? (Quel est le nombre moyen d'individus à la $n^{\text{ème}}$ génération ?)"

Théorème

On a : $\forall n \in \mathbb{N}, \mathbb{E}[Z_n] = m^n$.

En effet, soit $n \in \mathbb{N}$;

$$\begin{aligned}
\mathbb{E}[Z_{n+1}] &= \mathbb{E}\left[\sum_{i=1}^{Z_n} X_{i,n}\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{Z_n} X_{i,n}\middle| Z_n\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{\infty} \mathbb{1}_{i\leqslant Z_n} X_{i,n}\middle| Z_n\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{\infty} \mathbb{1}_{i\leqslant Z_n} \mathbb{E}[X_{i,n}|Z_n]\right] \text{ (par Fubini-Tonelli, car } \mathbb{1}_{i\leqslant Z_n} X_{i,n} \geqslant 0 \text{ ps et par } Z_n\text{-mesurabilité de } \mathbb{1}_{i\leqslant Z_n}) \\
&= \mathbb{E}\left[\sum_{i=1}^{Z_n} \mathbb{E}[X_{i,n}]\right] \text{ (d'après le lemme, on a l'indépendance entre } X_{i,n} \text{ et } Z_n) \\
&= \mathbb{E}\left[\sum_{i=1}^{Z_n} m\right] = m\mathbb{E}[Z_n]
\end{aligned}$$

On conclut par récurrence, en utilisant le fait que $Z_0 = 1$.

Proposition

On définit la série génératrice de X par $G : s \mapsto \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} p_k s^k$.

On a les résultats suivants :

1. G est bien définie sur $[0,1]$ et y est de classe $\mathcal{C}^1$.
2. (a) G est strictement croissante sur $]0,1[$.
   (b) G est convexe sur $]0,1[$.
   (c) G est strictement convexe sur $]0,1[\Leftrightarrow p_0 + p_1 < 1$.

Démonstration :

1. $\forall k \in \mathbb{N}, s \mapsto p_k s^k$ est de classe $\mathcal{C}^1$ sur $[0,1]$, la série $\sum_{k \geqslant 0} p_k 1^k$ converge (vers 1), et la série de fonctions $\sum_{k \geqslant 1} k p_k s^{k-1}$ converge normalement (car X est intégrable) donc uniformément sur $[0,1]$.
   Par conséquent, la série $\sum_{k \geqslant 0} p_k s^k$ converge uniformément vers G, de classe $\mathcal{C}^1$ sur $[0,1]$.
2. La série entière $\sum_{k \geqslant 0} p_k s^k$ ayant un rayon de convergence $\geqslant 1$, on a :

$$\forall s \in [0,1[, G'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1} \text{ et } G''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2}$$

   Comme $p_0 < 1$, on a : $\exists k_0 > 0, p_{k_0} > 0$.
   (a) Ainsi : $\forall s \in]0,1[, G'(s) \geqslant k_0 p_{k_0} s^{k_0 - 1} > 0$ et G est strictement croissante sur $]0,1[$.
   (b) Aussi : $\forall s \in]0,1[, G''(s) \geqslant k_0 (k_0 - 1) p_{k_0} s^{k_0 - 2} \geqslant 0$ et G est convexe sur $]0,1[$.
   (c) Si $p_0 + p_1 = 1$, alors on a $k_0 = 1$ et G est affine donc n'est pas strictement convexe sur $]0,1[$.
   Si $p_0 + p_1 < 1$, alors on peut avoir $k_0 > 1$ et $G'' > 0$ sur $]0,1[$ d'où la stricte convexité. ■

Proposition

Pour $n \in \mathbb{N}$, on définit la série génératrice de Z_n par $G_n : s \mapsto \mathbb{E}\left[s^{Z_n}\right] = \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) s^k$.

Comme précédemment, on peut montrer que G_n est bien définie sur $[0,1]$.

On a : $\forall n \in \mathbb{N}^*, G_n = \underbrace{G \circ \ldots \circ G}_{n \text{ fois}}$ sur $[0,1]$.

Démonstration :

Soit $n \in \mathbb{N}, s \in [0,1]$,

$$G_{n+1}(s) = \mathbb{E}\left[s^{Z_{n+1}}\right] = \mathbb{E}\left[s^{\sum_{i=1}^{Z_n} X_{i,n}}\right] = \mathbb{E}\left[\prod_{i=1}^{Z_n} s^{X_{i,n}}\right] = \mathbb{E}\left[\sum_{j=0}^{\infty} \mathbb{1}_{Z_n = j} \prod_{i=1}^{j} s^{X_{i,n}}\right]$$

$$= \sum_{j=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{Z_n = j} \prod_{i=1}^{j} s^{X_{i,n}}\right] \text{(par Fubini-Tonelli, car les termes sont tous positifs)}$$

$$= \sum_{j=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{Z_n = j}\right] \prod_{i=1}^{j} \mathbb{E}\left[s^{X_{i,n}}\right] \text{(car les variables en jeu ici sont toutes indépendantes)}$$

$$= \sum_{j=0}^{\infty} \mathbb{P}(Z_n = j) \, \mathbb{E}\left[s^X\right]^j = \sum_{j=0}^{\infty} \mathbb{P}(Z_n = j) \, G(s)^j = G_n(G(s))$$

On conclut par récurrence, en utilisant le fait que $Z_1 = X_{0,0}$ suit la loi de X. ■

Proposition

La probabilité d'extinction π_∞ est le plus petit point fixe de G sur l'intervalle $[0,1]$.

Démonstration :

La proposition précédente donne : $\forall n \in \mathbb{N}, \forall s \in [0,1], G_{n+1}(s) = G(G_n(s))$.
En évaluant en 0, on obtient la relation : $\pi_{n+1} = G(\pi_n)$, puis par continuité de G sur $[0,1]$, on obtient que π_∞ est un point fixe de G. Reste à montrer que c'est le plus petit.
Soit $u \in [0,1]$ un point fixe de G. On va montrer par récurrence que $\forall n \in \mathbb{N}^*, \pi_n \leqslant u$.

- On a : $\pi_1 = G(\pi_0) = G(\mathbb{P}(Z_0 = 0)) = G(0) \leqslant G(u) = u$, car G est croissante.
- Si $\pi_n \leqslant u$, alors $\pi_{n+1} = G(\pi_n) \leqslant G(u) = u$, toujours par croissance de G.

Par conséquent, $\forall n \in \mathbb{N}, \pi_n \leqslant u$, puis par passage à la limite : $\pi_\infty \leqslant u$. ■

Théorème

Si $m \leqslant 1$, alors $\pi_\infty = 1$.
Si $m > 1$, alors π_∞ est l'unique point fixe de G sur $]0,1[$.

Démonstration :

On rappelle qu'on a deux cas :

- Si $p_0 + p_1 = 1$, et alors G est une fonction affine ; comme $p_0 > 0$, la droite représentative de G coupe en un unique point la droite d'équation $y = x$. Nécessairement, ce point d'intersection a pour coordonnées $(1,1)$.
- Sinon, G est strictement convexe sur $]0,1[$; il en va alors de même pour $x \mapsto G(x) - x$ qui s'annule donc au plus deux fois [2].

Dans tous les cas, on a : $G(x) - x$ s'annule au plus 2 fois sur $[0,1]$; aussi $G'(0) = p_1$ et $G'(1) = \sum_{n=1}^{\infty} n p_n = m$.

Supposons $m > 1$.

Alors $G' - 1$ est une fonction croissante de $p_1 - 1 < 0$ (car $p_0 > 0$) à $m - 1 > 0$, donc elle s'annule en un point $\alpha \in]0,1[$.
La fonction $G-\mathrm{Id}$ est alors décroissante sur $[0,\alpha]$ puis croissante sur $[\alpha, 1]$.
Comme $G(0) - 0 = p_0 > 0$ et $G(1) - 1 = 0$, il existe un point dans l'intervalle $]0,\alpha]$ où $G-\mathrm{Id}$ s'annule.
π_∞ est donc l'unique point fixe de G sur l'intervalle $]0,1[$ (car G en a au plus 2).

x	0	π_∞	α		1
$G'(x)-1$	$p_1 - 1$	$-$	0	$+$	$m-1$
$G(x)-x$	p_0 $\searrow$	0 $\searrow$		$\nearrow$	0

Supposons $m \leqslant 1$.

Alors $G' - 1$ est une fonction croissante sur $[0,1]$, négative ou nulle en 1 ; donc négative sur $[0,1]$.
Donc $G-\mathrm{Id}$ est décroissante sur $[0,1]$, et s'annule en 1.
Comme cette fonction admet au plus 2 annulations, elle ne s'annule qu'en 1 (car sinon elle s'annulerait sur un intervalle non-réduit à un singleton).
Par conséquent, $\pi_\infty = 1$.

x	0		1
$G'(x)-1$	$p_1 - 1$	$-$	$m-1$
$G(x)-x$	p_0	$\searrow$	0

■

2. Effectivement, si cette fonction s'annule en trois points distincts, par le théorème de Rolle, sa dérivée s'annulera en deux points distincts, $a < b$. Mais $x \mapsto G(x) - x$ est convexe, donc sa dérivée est croissante, donc nulle sur $[a,b]$. Ceci implique que $G' - 1$ n'est pas strictement croissante, et donc que $x \mapsto G(x) - x$ n'est pas strictement convexe. Attention à l'idée reçue qui consiste à penser que "strictement convexe" équivaut à "dérivée seconde strictement positive" (pour les fonctions deux fois dérivables) – pour vous en convaincre, considérez $x \mapsto x^4$.

Inégalité de Hoeffding et application [1]

[Ouv2], exercice 10.11

Théorème

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes et centrées.
De plus, on suppose que $\forall n \in \mathbb{N}^*$, X_n est ps bornée par c_n, où $c_n > 0$.
On note : $\forall n \in \mathbb{N}^*, S_n = \sum_{j=1}^{n} X_j$ et $a_n = \sum_{j=1}^{n} c_j^2$.
Alors $\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.

Démonstration :

→ Il va s'agir de démontrer le lemme qui suit.

Lemme

Soit X une variable aléatoire réelle, centrée, et ps bornée par 1.
On note L_X sa transformée de Laplace.
On a : $\forall t \in \mathbb{R}, L_X(t) \leqslant \exp\left(\frac{t^2}{2}\right)$.

Démonstration :

Soit $t \in \mathbb{R}$, on a : $\forall x \in [-1,1], tx = \frac{1-x}{2}(-t) + \frac{1+x}{2}t$.
Donc, par convexité de l'exponentielle, on en déduit : $\forall x \in [-1,1], \mathrm{e}^{tx} \leqslant \frac{1-x}{2}\mathrm{e}^{-t} + \frac{1+x}{2}\mathrm{e}^{t}$.
Mais on sait que $|X| \leqslant 1$ ps ; en particulier, e^{tX} est bornée ps donc admet un moment d'ordre 1.
Ainsi, L_X est bien définie en t et :

$$L_X(t) \leqslant \frac{\mathbb{E}[1-X]}{2}\mathrm{e}^{-t} + \frac{\mathbb{E}[1+X]}{2}\mathrm{e}^{t} = \frac{1}{2}\left(\mathrm{e}^{-t} + \mathrm{e}^{t}\right) = \operatorname{ch} t = \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} \leqslant \sum_{n=0}^{\infty} \frac{t^{2n}}{n!2^n} = \exp\left(\frac{t^2}{2}\right).$$ ■

→ Fixons $n \in \mathbb{N}^*$.
On applique le lemme aux variables $\frac{X_j}{c_j}$, où $j \in [\![1,n]\!]$: $\forall t \in \mathbb{R}, L_{X_j}(t) = L_{\frac{X_j}{c_j}}(tc_j) \leqslant \exp\left(\frac{t^2}{2}c_j^2\right)$.
Mais par indépendance des $\exp(tX_j)$ pour $j \in [\![1,n]\!]$, il vient :

$$\forall t \in \mathbb{R}, L_{S_n}(t) = \prod_{j=1}^{n} L_{X_j}(t) \leqslant \exp\left(\frac{t^2}{2}a_n\right).$$

→ Soient $t > 0$ et $\varepsilon > 0$; on a : $S_n > \varepsilon \Leftrightarrow \mathrm{e}^{tS_n} > \mathrm{e}^{t\varepsilon}$.
Ainsi, par l'inégalité de Markov, on obtient :

$$\mathbb{P}(S_n > \varepsilon) = \mathbb{P}\left(\mathrm{e}^{tS_n} > \mathrm{e}^{t\varepsilon}\right) \leqslant \frac{\mathbb{E}\left[\mathrm{e}^{tS_n}\right]}{\mathrm{e}^{t\varepsilon}} = \mathrm{e}^{-t\varepsilon}L_{S_n}(t) \leqslant \exp\left(-t\varepsilon + \frac{t^2}{2}a_n\right).$$

Cette inégalité est vraie pour tout $t > 0$, donc en particulier pour $t = \frac{\varepsilon}{a_n}$, où $-t\varepsilon + \frac{t^2}{2}a_n$ réalise son minimum, d'où : $\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(\frac{\varepsilon^2}{a_n^2}\frac{a_n}{2} - \frac{\varepsilon}{a_n}\varepsilon\right) = \exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.

→ Soit $\varepsilon > 0$ quelconque.
On a : $\mathbb{P}(|S_n| > \varepsilon) = \mathbb{P}(S_n > \varepsilon) + \mathbb{P}(S_n < -\varepsilon)$.
Mais on aurait pu appliquer tout ce qu'on vient de faire aux variables $-X_j$, où $j \in [\![1,n]\!]$, et donc :
$\mathbb{P}(S_n < -\varepsilon) = \mathbb{P}(-S_n > \varepsilon) \leqslant \exp\left(-\frac{\varepsilon^2}{2a_n}\right)$.
Et finalement, $\mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2a_n}\right)$. ■

Corollaire

Soit $\alpha > 0$; on ajoute l'hypothèse supplémentaire : $\exists \beta > 0, \forall n \in \mathbb{N}^*, a_n \leqslant n^{2\alpha-\beta}$.
Alors : $\dfrac{S_n}{n^\alpha} \xrightarrow[n\to\infty]{\text{ps}} 0$.

Démonstration :

Soit $\varepsilon > 0$, on a, par Hoeffding : $\forall n \in \mathbb{N}^*, \mathbb{P}\left(|S_n| > n^\alpha \varepsilon\right) \leqslant 2\exp\left(-\dfrac{n^{2\alpha}\varepsilon^2}{2a_n}\right) \leqslant 2\exp\left(-\dfrac{n^\beta\varepsilon^2}{2}\right)$.

Mais la série $\sum_{n\geqslant 1} \exp\left(-\dfrac{n^\beta\varepsilon^2}{2}\right)$ converge (par le critère de Riemann), car $\exists N \in \mathbb{N}^*, \forall n \leqslant N, \dfrac{\varepsilon^2}{2}n^\beta \geqslant 2\ln n$ et donc $\exists N \in \mathbb{N}^*, \forall n \leqslant N, 0 \leqslant \exp\left(-\dfrac{n^\beta\varepsilon^2}{2}\right) \geqslant \dfrac{1}{n^2}$.

Ainsi, la série $\sum_{n\geqslant 1} \mathbb{P}\left(|S_n| > n^\alpha\varepsilon\right)$ converge, d'où, par Borel-Cantelli :

$$\forall \varepsilon > 0, \mathbb{P}\left(\limsup_{n\to\infty}\{|S_n| > n^\alpha\varepsilon\}\right) = 0, \text{ ie } \mathbb{P}\left(\liminf_{n\to\infty}\{|S_n| \leqslant n^\alpha\varepsilon\}\right) = 1.$$

En particulier, $\forall p \in \mathbb{N}^*, \mathbb{P}\left(\bigcup_{n\in\mathbb{N}^*}\bigcap_{k\geqslant n}\left\{\left|\dfrac{S_k}{k^\alpha}\right| \leqslant \dfrac{1}{p}\right\}\right) = 1$.

C'est-à-dire : $\forall p \in \mathbb{N}^*, \exists N_p$ négligeable, $\forall \omega \in N_p^c, \exists n \in \mathbb{N}, \forall k \geqslant n, \left|\dfrac{S_k(\omega)}{k^\alpha}\right| \leqslant \dfrac{1}{p}$.

On pose alors $N = \bigcup_{p\in\mathbb{N}^*} N_p$, alors N est négligeable et :

$$\forall \omega \in N^c, \forall p \in \mathbb{N}^*, \exists n \in \mathbb{N}, \forall k \geqslant n, \left|\frac{S_k(\omega)}{k^\alpha}\right| \leqslant \frac{1}{p}.$$

Ou, en d'autres termes : $\dfrac{S_n}{n^\alpha} \xrightarrow[n\to\infty]{\text{ps}} 0$. ∎

Références

[Ouv2] J.-Y. OUVRARD – *Probabilités 2*, 3e éd., Cassini, 2009.

1. Une autre application, en statistiques ; elle provient de la partie 3.2 du livre *Statistique mathématique*, de B. CADRE et C. VIAL, paru en 2012 aux éditions Ellipses. Plaçons-nous dans un modèle statistique $(\mathcal{H}^n, \{P_\theta\}_{\theta\in\Theta})$ avec $\mathcal{H} \subset \mathbb{R}$ et $\Theta \subset \mathbb{R}^d$. Le paramètre d'intérêt est $g(\theta)$ avec $g : \Theta \to \mathbb{R}$ une fonction continue. On suppose que $X_1, \dots, X_n$ sont indépendantes et identiquement distribuées, de loi P_θ, bornées P_θ-ps et avec $\mathbb{E}_\theta[X_1] = g(\theta)$. Soit c une borne P_θ-presque sûre de $X_1 - g(\theta)$. On veut un intervalle de confiance pour $g(\theta)$.

Par Hoeffding, $P_\theta\left(|\overline{X_n} - g(\theta)| > \varepsilon\right) = P_\theta\left(\left|\sum_{j=1}^n \left(X_j - g(\theta)\right)\right| > n\varepsilon\right) \leqslant 2\exp\left(-\dfrac{n^2\varepsilon^2}{2nc^2}\right) = 2\exp\left(-\dfrac{n\varepsilon^2}{2c^2}\right)$. Soit $\alpha \in]0,1[$, on choisit ε de sorte que $\alpha = 2\exp\left(-\dfrac{n\varepsilon^2}{2c^2}\right)$, ie : $\varepsilon = c\sqrt{\dfrac{2}{n}\ln\dfrac{2}{\alpha}}$. Dès lors, on obtient l'intervalle de confiance par excès au niveau $1-\alpha$, pour $g(\theta)$: $I_\alpha = \left[\overline{X_n} - c\sqrt{\dfrac{2}{n}\ln\dfrac{2}{\alpha}}, \overline{X_n} + c\sqrt{\dfrac{2}{n}\ln\dfrac{2}{\alpha}}\right]$.