Compléments de probabilités et statistique
Notes du Cours 2021

Anton Thalmaier

Unité de Recherche en Mathématiques
Université du Luxembourg
Campus Belval – Maison du Nombre
L–4364 Esch-sur-Alzette

Email address: anton.thalmaier@uni.lu
URL: math.uni.lu/thalmaier

Table des matières

CHAPTER 1

Marches aléatoires

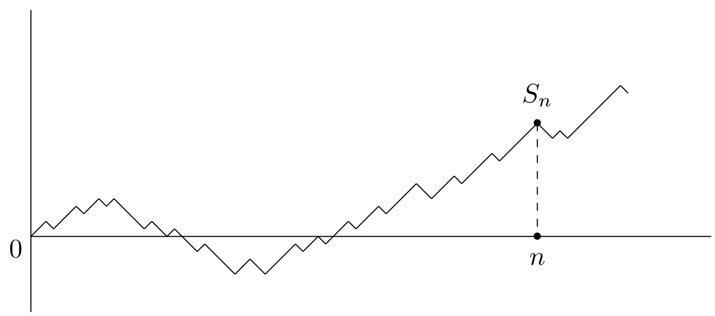
1.1. Marche aléatoire symétrique

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et Bernoulli telle que

$$\mathbb{P}\{X_n = 1\} = \mathbb{P}\{X_n = -1\} = \frac{1}{2}.$$

On appelle *marche aléatoire symétrique* sur $\mathbb{Z}$ la suite $(S_n)_{n\geqslant 0}$:

$$\begin{cases} S_0 = 0 \\ S_n = X_1 + \ldots + X_n, \quad n \geqslant 1. \end{cases}$$



Interprétation. Un joueur joue à pile ou face (avec une pièce équilibré). Il reçoit 1 euro s'il obtient pile, et donne 1 euro à la banque s'il obtient face. Alors S_n est la fortune du joueur après le $n^{\text{ième}}$ coup.

On remarque que

$$X_n = S_n - S_{n-1} = \begin{cases} +1 & \text{pour pile} \;\; \text{au } n^{\text{ième}} \text{ coup}, \\ -1 & \text{pour face} \;\; \text{au } n^{\text{ième}} \text{ coup}. \end{cases}$$

Le vecteur $(S_0, S_1, \ldots, S_N)$ prends des valeurs dans $\mathbb{Z}^{N+1}$ telles que

$$S_0 = 0 \quad \text{et} \quad S_i - S_{i-1} \in \{\pm 1\}, \quad i = 1, \ldots, N.$$

Pour $(s_0, s_1, \ldots, s_N) \in \mathbb{Z}^N$ avec $s_0 = 0$ et $s_i - s_{i-1} \in \{\pm 1\}$, $i = 1, \ldots, N$, on trouve

$$\begin{aligned} &\mathbb{P}\{S_0 = s_0,\ S_1 = s_1, \ldots,\ S_N = s_N\} \\ &\qquad = \mathbb{P}\{X_1 = s_1 - s_0,\ X_2 = s_2 - s_1, \ldots, X_N = s_N - s_{N-1}\} \\ &\qquad = \prod_{i=1}^{N} \underbrace{\mathbb{P}\{X_i = s_i - s_{i-1}\}}_{=1/2} = \frac{1}{2^N}. \end{aligned}$$

REMARQUE 1.1. Évidement, on a

$$\mathbb{E}[S_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = 0, \quad \text{car } \mathbb{E}[X_1] = \ldots = \mathbb{E}[X_n] = 0,$$

et

$$\mathbb{E}[S_n^2] = \sum_{i,j=1}^{n} \mathbb{E}[X_i X_j] = n, \quad \text{car } \mathbb{E}[X_i X_j] = \begin{cases} 1, & i = j \\ \mathbb{E}[X_i]\,\mathbb{E}[X_j] = 0, & i \neq j. \end{cases}$$

Problème. Le joueur se demande s'il peut améliorer son gain en arrêtant de jouer à un instant favorable.

NOTATION 1.2. Soit $(S_n)_{n\geq 0}$ une suite de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$. Pour $n \in \mathbb{N}$, on note

$$\mathscr{A}_n = \sigma\{S_0, \ldots, S_n\}$$

la tribu engendrée par $S_0, \ldots, S_n$. Alors:

$$A \in \mathscr{A}_n \iff \begin{cases} A \text{ est réunion des événements de type} \\ \{S_0 = s_0, S_1 = s_1, \ldots, S_n = s_n\}, \quad s_i \in \mathbb{Z}. \end{cases}$$

On a

$$\{\varnothing, \Omega\} = \mathscr{A}_0 \subset \mathscr{A}_1 \subset \mathscr{A}_2 \subset \ldots \subset \mathscr{A}.$$

La tribu $\mathscr{A}_n$ s'appelle "*tribu des événements disponible à l'instant* n".

DÉFINITION 1.3. On appelle *temps d'arrêt* une variable aléatoire

$$T\colon \Omega \to \mathbb{N} \cup \{\infty\}$$

telle que

$$\{T = n\} \in \mathscr{A}_n, \quad \forall\, n \in \mathbb{N}.$$

REMARQUE 1.4. Pour un temps d'arrêt T on a

$$\forall\, n \in \mathbb{N},\ \{T = n\} \in \mathscr{A}_n \iff \forall\, n \in \mathbb{N},\ \{T \leqslant n\} \in \mathscr{A}_n.$$

PREUVE. Il faut de montrer les deux directions.

"$\Rightarrow$": Pour tout $n \in \mathbb{N}$, on a

$$\{T \leqslant n\} = \bigcup_{k=0}^{n} \{T = k\} \in \mathscr{A}_n,$$

car $\{T = k\} \in \mathscr{A}_k \subset \mathscr{A}_n$.

"$\Leftarrow$": Pour tout $n \in \mathbb{N}$,

$$\{T = n\} = \{T \leqslant n\} \setminus \{T \leqslant n-1\} \in \mathscr{A}_n,$$

car $\{T \leqslant n\} \in \mathscr{A}_n$ et $\{T \leqslant n-1\} \in \mathscr{A}_{n-1} \subset \mathscr{A}_n$. $\square$

PROPOSITION 1.5. *Soient R et T deux temps d'arrêt. Alors*

$$R \wedge T = \min(R, T) \quad R \vee T = \max(R, T), \quad R + T$$

sont aussi des temps d'arrêt.

PREUVE. 1. $R \wedge T$ est un temps d'arrêt. En effet,

$$\forall\, n \in \mathbb{N}, \quad \{R \wedge T \leqslant n\} = \underbrace{\{R \leqslant n\}}_{\in \mathscr{A}_n} \cup \underbrace{\{T \leqslant n\}}_{\in \mathscr{A}_n} \in \mathscr{A}_n.$$

2. $R \vee T$ temps d'arrêt, car
$$\forall\, n \in \mathbb{N}, \quad \{R \vee T \leqslant n\} = \underbrace{\{R \leqslant n\}}_{\in \mathscr{A}_n} \cap \underbrace{\{T \leqslant n\}}_{\in \mathscr{A}_n} \in \mathscr{A}_n.$$

3. R+T temps d'arrêt, car
$$\forall\, n \in \mathbb{N}, \quad \{R + T \leqslant n\} = \bigcup_{k=0}^{n} \underbrace{\{R = k\}}_{\in \mathscr{A}_k \subset \mathscr{A}_n} \cap \underbrace{\{T = n-k\}}_{\in \mathscr{A}_{n-k} \subset \mathscr{A}_n} \in \mathscr{A}_n. \qquad \square$$

LEMME 1.6. *Pour tout $n \geqslant 0$ et tout $A_n \in \mathscr{A}_n$, on a*
$$\mathbb{E}[X_{n+1}\, \mathbb{1}_{A_n}] = 0.$$

PREUVE. Sans perte de généralité, on peut supposer que
$$A_n = \{S_0 = s_0,\ S_1 = s_1, \dots, S_n = s_n\} \neq \varnothing.$$
Alors
$$\begin{aligned} A_n &= \{X_1 = s_1 - s_0,\ X_2 = s_2 - s_1, \dots, X_n = s_n - s_{n-1}\} \\ &= \{X_1 = x_1, \dots, X_n = x_n\}, \end{aligned}$$
où $x_i = s_i - s_{i-1} \in \{\pm 1\}$, et donc
$$\mathbb{P}\big(A_n \cap \{X_{n+1} = 1\}\big) = \frac{1}{2^{n+1}} = \frac{1}{2}\, \mathbb{P}(A_n).$$
En particulier,
$$\begin{aligned} \mathbb{E}[X_{n+1}\, \mathbb{1}_{A_n}] &= \mathbb{P}\big(A_n \cap \{X_{n+1} = 1\}\big) - \mathbb{P}\big(A_n \cap \{X_{n+1} = -1\}\big) + 0 \cdot \mathbb{P}\big(A_n^c\big) \\ &= \frac{1}{2}\, \mathbb{P}(A_n) - \frac{1}{2}\, \mathbb{P}(A_n) = 0. \end{aligned} \qquad \square$$

PROPOSITION 1.7. *Soit $N \in \mathbb{N}$ et soit T un temps d'arrêt tel que $T \leqslant N$. On a*
$$\mathbb{E}[S_T] = 0.$$

PREUVE. Par définition, on a $(S_T)(\omega) = S_{T(\omega)}(\omega)$ et
$$S_T = \sum_{n=1}^{T} X_n = \sum_{n=1}^{N} X_n\, \mathbb{1}_{\{n \leqslant T\}}.$$
Mais $\{T \geqslant n\} = \{T \leqslant n-1\}^c \in \mathscr{A}_{n-1}$ et donc
$$\mathbb{E}[S_T] = \sum_{n=1}^{N} \mathbb{E}[X_n\, \mathbb{1}_{\{n \leqslant T\}}] = 0, \quad \text{par Lemme 1.6.} \qquad \square$$

REMARQUE 1.8. On a même:
$$\mathbb{E}[S_T] = 0, \quad \forall \text{ temps d'arrêt } T \text{ tel que } \mathbb{E}[T] < \infty.$$

PREUVE. D'abord on remarque que
$$\mathbb{E}[T] < \infty \Rightarrow \mathbb{P}\{T < \infty\} = 1 \Rightarrow S_T \text{ bien définie.}$$
Comme $S_T = \sum\limits_{n=1}^{\infty} X_n\, \mathbb{1}_{\{n \leqslant T\}}$, on a
$$|S_T| \leqslant \sum_{n=1}^{\infty} \underbrace{|X_n|}_{=1}\, \mathbb{1}_{\{n \leqslant T\}} = T,$$

et de même façon, $|S_{T\wedge N}| \leqslant T \wedge N \leqslant T$. Par convergence dominée,

$$\mathbb{E}[S_T] = \lim_{N\to\infty} \underbrace{\mathbb{E}[S_{T\wedge N}]}_{=0} = 0.$$

En effet,

$$\begin{aligned}\big|\mathbb{E}[S_T] - \mathbb{E}[S_{T\wedge N}]\big| &\leqslant \mathbb{E}\left[\sum_{n=1}^{\infty} X_n \,\mathbb{1}_{\{T\wedge N < n \leqslant T\}}\right] \\ &\leqslant \mathbb{E}[T - T\wedge N] \downarrow 0, \quad N \to \infty.\end{aligned}$$

□

Exemple 1.9. Pour $a \in \mathbb{Z}$, soit

$$T_a = \inf\{n \geqslant 0 \mid S_n = a\} \quad (\inf \varnothing = \infty).$$

Alors T_a est un temps d'arrêt. En effet, pour $a \neq 0$ (sans restrictions), on a

$$\{S_k = a\} = \bigcup_{\substack{x_1,\ldots,x_k \in \{\pm 1\} \\ x_1+\ldots+x_k = a}} \{X_1 = x_1, \ldots, X_k = x_k\} \in \mathscr{A}_k.$$

Donc, pour tout $n \in \mathbb{N}$,

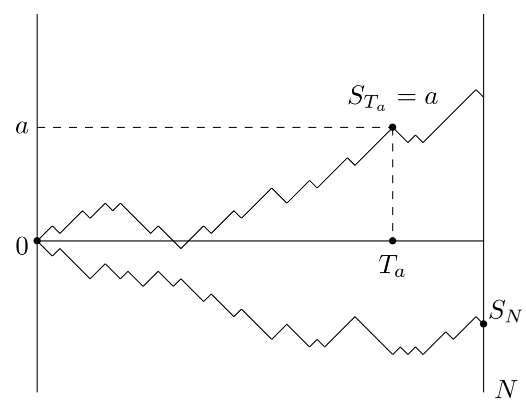
$$\{T_n \leqslant n\} = \bigcup_{k=1}^{n} \underbrace{\{S_k = a\}}_{\in \mathscr{A}_k \subset \mathscr{A}_n} \in \mathscr{A}_n.$$

En particulier,

$$T_{a,N} := T_a \wedge N$$

est un temps d'arrêt pour tout $N \in \mathbb{N}$ et on a

$$\begin{aligned}0 = \mathbb{E}[S_{T_{a\wedge N}}] &= \mathbb{E}[S_{T_a} \,\mathbb{1}_{\{T_a \leqslant N\}}] + \mathbb{E}[S_N \,\mathbb{1}_{\{T_a > N\}}] \\ &= a\,\mathbb{P}\{T_a \leqslant N\} + \mathbb{E}[S_N \,\mathbb{1}_{\{T_a > N\}}].\end{aligned}$$



Proposition 1.10. *Pour tout $a \neq 0$, on a $\mathbb{E}[T_a] = \infty$.*

Preuve. Supposons que $\mathbb{E}[T_a] < \infty$, alors $\mathbb{P}\{T_a < \infty\} = 1$, c.à.d. $T_a < \infty$ presque sûrement. Donc

$$0 = \mathbb{E}[S_{T_a}] = a, \quad \text{car } S_{T_a} = a,$$

ce qui donne une contradiction si $a \neq 0$. □

Resumé 1.11 (Marche aléatoire). Soit

$$S_n := \begin{cases} 0, & \text{si } n = 0, \\ X_1 + \ldots + X_n, & \text{si } n \geqslant 1, \end{cases}$$

une marche aléatoire symétrique et

$$\mathscr{A}_n = \sigma\{S_0, S_1, \ldots, S_n\} = \sigma\{X_1, \ldots, X_n\}.$$

Si $T\colon \Omega \to \overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ est un temps d'arrêt, c.à.d. $\{T \leqslant n\} \in \mathscr{A}_n$ pour tout n, alors $\mathbb{E}[S_T] = 0$ si $\mathbb{E}[T] < \infty$. En particular, si on note

$$T_a = \inf\{n \geqslant 0 \mid S_n = a\} \quad \text{et} \quad T_{a,N} = T_a \wedge N = \min(T_a, N), \quad a \in \mathbb{Z},$$

alors on a $\mathbb{E}[S_{T_{a,N}}] = 0$, mais $\mathbb{E}[T_a] = \infty$ pour $a \neq 0$.

1.2. Stratégie de jeu et processus arrêtés

Définition 1.12 (Stratégie de jeu). Soit $\beta = (\beta_k)_{k \geqslant 1}$ une suite *prévisible* de variables bornées, c.à.d. β_k est $\mathscr{A}_{k-1}$ mesurable pour tout $k \geqslant 1$ au sens que

$$\forall\, b \in \mathbb{R}, \quad \{\beta_k \leqslant b\} \in \mathscr{A}_{k-1}.$$

On appelle $(\beta_k)_{k \geqslant 1}$ une *stratégie* de jeu; β_k représentant la somme misée au $k^{\text{ième}}$ coup. On pose

$$S_n^{(\beta)} = \sum_{k=1}^{n} \beta_k X_k = \sum_{k=1}^{n} \beta_k (S_k - S_{k-1}); \qquad S_0^{(\beta)} := 0.$$

La variable $S_n^{(\beta)}$ correspond au gain après n coups.

Proposition 1.13. *Les suites suivantes satisfont les hypothèses d'une stratégie*:

(1) $\beta_k = \mathbb{1}_{\{T \geqslant k\}}$ *avec* T *est un temps d'arrêt;*
(2) *(Doublement successif jusqu'au premier* $+1$*)*

$$\beta_k = 2^{k-1}\, \mathbb{1}_{\{T \geqslant k\}} \quad \textit{avec} \quad T := \inf\{n \geqslant 1 \mid X_n = +1\};$$

(3) $\beta_k = S_{k-1}\, \mathbb{1}_{\{T \geqslant k\}}$ *avec* T *un temps d'arrêt.*

Preuve.

(1) Soit $\beta_k = \mathbb{1}_{\{T \geqslant k\}}$. Comme T est un temps d'arrêt, on a

$$\{T \geqslant k\} = \{T \leqslant k-1\}^c \in \mathscr{A}_{k-1} \quad \forall\, k,$$

et puis

$$S_n^{(\beta)} = \sum_{k=1}^{n} \mathbb{1}_{\{T \geqslant k\}} X_k = \sum_{k=1}^{n \wedge T} X_k = S_{n \wedge T}.$$

On remarque que par définition, $(S_{n \wedge T})(\omega) = (S_{n \wedge T(\omega)})(\omega)$.

(2) Soit $\beta_k = 2^{k-1}\, \mathbb{1}_{\{T \geqslant k\}}$ et $T = \inf\{n \geqslant 1 \,|\, X_n = 1\}$. Comme T est un temps d'arrêt, on a

$$\{T \geqslant k\} \in \mathscr{A}_{k-1} \quad \forall\, k.$$

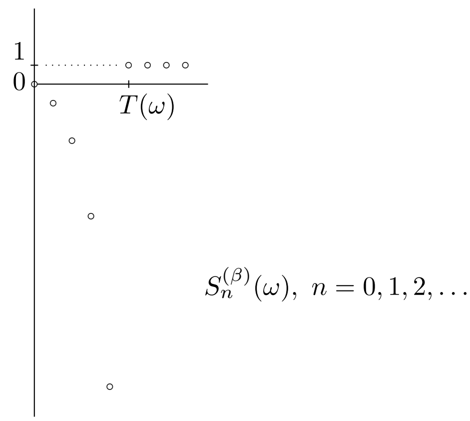
Évidement,

$$\beta_1 = 1 \quad \text{et} \quad \beta_k(\omega) = \begin{cases} 2^{k-1} & \text{si } X_1(\omega) = \ldots = X_{k-1}(\omega) = -1 \\ 0 & \text{sinon.} \end{cases}$$

Par exemple, si $T(\omega) = n$, alors

$$S_{n-1}(\omega) = -\sum_{k=1}^{n-1} 2^{k-1} = -\sum_{k=0}^{n-2} 2^k = -(2^{n-1}-1), \quad \text{mais}$$
$$S_n(\omega) = S_{n-1}(\omega) + \beta_n = -(2^{n-1}-1) + 2^{n-1} = 1.$$

Pour la première égalité on utilise que $\sum_{k=0}^{N-1} 2^k = 2^N - 1$.



(3) Soit $\beta_k = S_{k-1}\,\mathbb{1}_{\{T \geqslant k\}}$. On observe que S_{k-1} et $\mathbb{1}_{\{T\geqslant k\}}$ sont $\mathscr{A}_{k-1}$-mesurable. En plus, comme

$$S_k^2 = (S_{k-1} + X_k)^2 = S_{k-1}^2 + 2S_{k-1}X_k + 1,$$

on a

$$\beta_k X_k = \frac{1}{2}(S_k^2 - S_{k-1}^2 - 1)\ \mathbb{1}_{\{T\geqslant k\}}.$$

Par conséquent, on obtient

$$S_n^{(\beta)} = \sum_{k=1}^n \beta_k X_k = \sum_{k=1}^{n\wedge T} \frac{1}{2}(S_k^2 - S_{k-1}^2 - 1) = \frac{1}{2}(S_{n\wedge T}^2 - n \wedge T). \qquad \square$$

Proposition 1.14. *Pour toute stratégie de jeu $(\beta_k)_{k\geqslant 1}$, on a*

$$\mathbb{E}[S_n^{(\beta)}] = 0, \quad n \geqslant 0.$$

Preuve. On a

$$\mathbb{E}[S_n^{(\beta)}] = \sum_{k=1}^n \mathbb{E}[\beta_k X_k],$$

donc il suffit à démontrer que

$$\mathbb{E}[\beta_k X_k] = 0, \quad \forall\, k \geqslant 1.$$

Comme β_k est $\mathscr{A}_{k-1}$-mesurable et comme $\operatorname{card}(\mathscr{A}_{k-1}) < \infty$, on peut écrire

$$\beta_k = \sum_{i=1}^{\ell} b_i\,\mathbb{1}_{\{\beta_k = b_i\}}, \quad \text{avec } b_1, \dots, b_\ell \in \mathbb{R}.$$

Par conséquent, on obtient par Lemme 1.6,

$$\mathbb{E}[\beta_k X_k] = \sum_{i=1}^{\ell} b_i \underbrace{\mathbb{E}[X_k \mathbb{1}_{\{\beta_k = b_i\}}]}_{=0} = 0,$$

car $\{\beta_k = b_i\} \in \mathscr{A}_{k-1}$. □

Corollaire 1.15.

(1) *Soit* $\beta_k = \mathbb{1}_{\{T \geqslant k\}}$. *Alors*

$$0 = \mathbb{E}[S_n^{(\beta)}] = \mathbb{E}[S_{T\wedge n}] \quad \xrightarrow[n\to\infty]{} \quad \mathbb{E}[S_T], \quad si \ \ \mathbb{E}[T] < \infty.$$

(2) *Soit* $\beta_k = 2^{k-1} \mathbb{1}_{\{T \geqslant k\}}$. *Alors*

$$0 = \mathbb{E}[S_n^{(\beta)}] = \frac{1}{2}\mathbb{E}[S_{n\wedge T}^2 - n \wedge T],$$

c.à.d.

$$\mathbb{E}[S_{n\wedge T}^2] = \mathbb{E}[n \wedge T]$$

et alors, quand $n \to \infty$,

$$\mathbb{E}[S_T^2] = \mathbb{E}[T], \quad si \ T \ est \ intégrable.$$

(égalité de Wald).

1.3. Premier temps de passage

Soit

$$S_n = \begin{cases} 0, & n = 0 \\ X_1 + \ldots + X_n, & n \geqslant 1 \end{cases}$$

avec (X_n) i.i.d. et $\mathbb{P}\{X_n = 1\} = \mathbb{P}\{X_n = -1\} = \frac{1}{2}$.

Lemme 1.16. *Pour tout* $a_1 < a_2$ *dans* $\mathbb{Z}$, *on a*

$$\mathbb{P}\{a_1 \leqslant S_n \leqslant a_2\} \to 0, \quad quand \ n \to +\infty.$$

Preuve. On note

$$U_n = \sum_{k=1}^{n} \mathbb{1}_{\{X_k = 1\}},$$

et donc

$$S_n = U_n - (n - U_n) = 2U_n - n.$$

On a

$$\mathbb{P}\{U_n = k\} = \binom{n}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k}\frac{1}{2^n}.$$

On remarque que

$$\begin{cases} n \text{ pair} & \Longrightarrow S_n \text{ pair}, \\ n \text{ impair} & \Longrightarrow S_n \text{ impair}, \end{cases}$$

c.à.d.

$$S_n \in \{-n, -n+2, \ldots, n-2, n\}.$$

Pour $k = 0, 1, \ldots, n$ on a:

$$\mathbb{P}\{S_n = 2k - n\} = \mathbb{P}\{2U_n - n = 2k - n\} = \mathbb{P}\{U_n = k\} = \binom{n}{k}\frac{1}{2^n}.$$

Comme

$$\binom{n}{k} \underset{[\geqslant]}{\leqslant} \binom{n}{k-1} \iff k \underset{[\geqslant]}{\leqslant} \frac{n+1}{2},$$

on obtient

$$\max_{\ell} \mathbb{P}\{S_n = \ell\} = \begin{cases} \mathbb{P}\{S_n = 0\} & n \text{ pair} \\ \mathbb{P}\{S_n = \pm 1\} & n \text{ impair.} \end{cases}$$

Par la formule de Stirling on a

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n},$$

d'où on obtient

$$\mathbb{P}\{S_{2n} = 0\} = \mathbb{P}\{S_{2n-1} = 1\} = \binom{2n}{n} \frac{1}{2^{2n}} = \frac{(2n)!}{n!n!} \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}} \to 0, \quad n \to \infty.$$

Donc, on a

$$\begin{aligned} \mathbb{P}\{a_1 \leqslant S_n \leqslant a_2\} &= \sum_{k=a_1}^{a_2} \mathbb{P}\{S_n = k\} \\ &\leqslant (a_2 - a_1 + 1)\, \mathbb{P}\{S_{2m} = 0\} \to 0, \quad n \to \infty, \end{aligned}$$

avec m tel que $n = 2m$ ou $n = 2m - 1$. □

Preuve du Lemme 1.16 (avec le théorème limite central). On utilise que

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}} = \frac{S_n}{\sqrt{n}} \xrightarrow{\text{en loi}} N(0,1), \quad \text{quand } n \to \infty.$$

Donc, pour n grand,

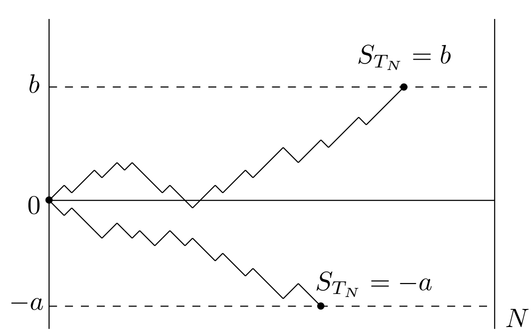
$$\mathbb{P}\{a_1 \leqslant S_n \leqslant a_2\} = \mathbb{P}\left\{\frac{a_1}{\sqrt{n}} \leqslant \frac{S_n}{\sqrt{n}} \leqslant \frac{a_2}{\sqrt{n}}\right\} \approx N(\frac{a_2}{\sqrt{n}}) - N(\frac{a_1}{\sqrt{n}}),$$

et par conséquent,

$$\mathbb{P}\{a_1 \leqslant S_n \leqslant a_2\} \to 0, \quad \text{quand } n \to \infty.$$ □

Exemple 1.17. Soient $a, b \in \mathbb{N}$ avec $-a < 0 < b$ et soit

$$T_N := T_{-a} \wedge T_b \wedge N.$$



Alors on a

$$0 = \mathbb{E}[S_{T_N}] = -a \underbrace{\mathbb{P}\{T_{-a} < T_b \ et \ T_{-a} \leqslant N\}}_{=:\ r_N^A} + b \underbrace{\mathbb{P}\{T_b < T_{-a} \ et \ T_b \leqslant N\}}_{=:\ r_N^A}$$
$$+ \underbrace{\mathbb{E}[S_N \, \mathbb{1}_{\{T_{-a} \wedge T_b > N\}}]}_{=:\ (*)}.$$

Quand $N \to \infty$, on obtient
(1) $r_N^A \to r^A := \mathbb{P}\{T_{-a} < T_b\}$
(2) $r_N^B \to r^B := \mathbb{P}\{T_b < T_{-a}\}$
(3) $|(*)| \leqslant (a \vee b)\, \mathbb{P}\{-a \leqslant S_N \leqslant b\} \to 0$.
Par conséquent,

$$-ar^A + br^B = 0.$$

D'autre part,

$$1 - (r_N^A + r_N^B) = \mathbb{P}\{T_{-a} \wedge T_b > N\} = \mathbb{P}\{-a \leqslant S_N \leqslant b\} \to 0, \text{ quand } N \to \infty,$$

donc

$$r^A + r^B = 1,$$

et puis,

$$r^A = \frac{b}{a+b}, \quad r^B = \frac{a}{a+b}.$$

PROPOSITION 1.18. *Pour $-a < 0 < b$, on a*

$$\mathbb{E}[T_{-a} \wedge T_b] = ab.$$

Rappelons que $\mathbb{E}[T_a] = +\infty$ pour tout $a \neq 0$.

PREUVE. Pour $T_N = T_{-a} \wedge T_b \wedge N$ on a:

$$\underbrace{\mathbb{E}[T_N]}_{\to \mathbb{E}[T_{-a} \wedge T_b]} = \mathbb{E}[S_{T_N}^2] = \underbrace{a^2 r_N^A}_{\to a^2 r^A} + \underbrace{b^2 r_N^B}_{\to b^2 r^B} + \underbrace{\mathbb{E}\left[S_N^2 \, \mathbb{1}_{\{T_{-a} \wedge T_b > N\}}\right]}_{\to 0}$$

avec

$$r_N^A = \mathbb{P}\{T_{-a} < T_b \ et \ T_{-a} \leqslant N\} \to r^A = \mathbb{P}\{T_{-a} < T_b\} = \frac{b}{a+b},$$
$$r_N^B = \mathbb{P}\{T_b < T_{-a} \ et \ T_b \leqslant N\} \to r^B = \mathbb{P}\{T_b < T_{-a}\} = \frac{a}{a+b},$$
$$r^A + r^B = 1.$$

Par conséquent,

$$\mathbb{E}[T_{-a} \wedge T_b] = a^2 r^A + b^2 r^B = \frac{a^2 b}{a+b} \frac{ab^2}{a+b} = ab. \qquad \square$$

LEMME 1.19 (Le principe de réflexion de Désiré André). *Soient $A = (m, a)$ et $B = (n, b)$ deux points du plan $\mathbb{Z}^2$. On suppose que $0 \leqslant m < n$ et $a \geqslant 1$, $b \geqslant 1$.*

Alors le nombre de chemins allant de A à B qui touchent ou traversent l'axe horizontal est égal au nombre de chemins allant du point A à $B' = (n, -b)$.

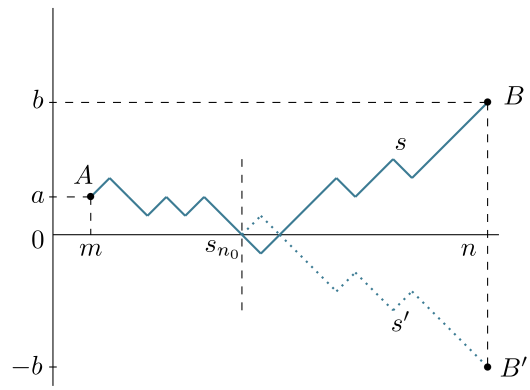
PREUVE. Soit

$$s := (s_m, s_{m+1}, \ldots, s_n)$$

un chemin de A à B, c.à.d.

$$s_m = a, \quad s_n = b, \quad s_{k+1} - s_k \in \{\pm 1\}, \quad k = m, \ldots, n-1,$$

touchant ou traversant l'axe horizontal.



Soit n_0 l'indice le plus à gauche où le chemin s atteint l'axe horizontal. Alors

$$s' = (s_m, \ldots, \underbrace{s_{n_0}}_{=0}, -s_{n_0+1}, \ldots, -s_n)$$

est un chemin de A à B' et l'application $s \mapsto s'$ est une bijection envoyant la première famille des chemins sur la seconde. □

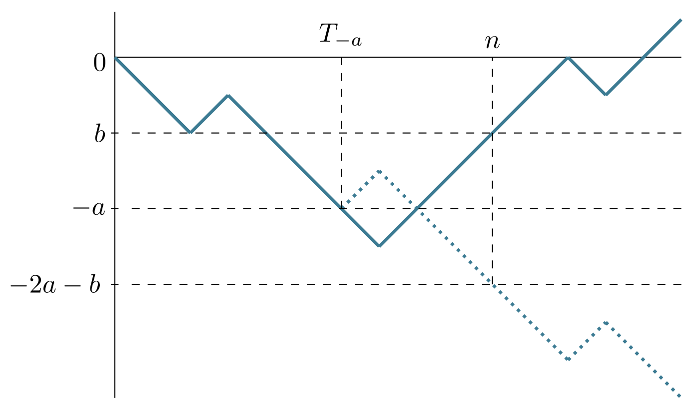
Notation. Pour $n \in \mathbb{N}$, on note $M_n = \min(S_0, S_1, \ldots, S_n)$.

COROLLAIRE 1.20 (Le principe de réflexion pour les marches aléatoires symétrique). *Soit $a > 0$ et $b \geqslant -a$. Alors:*

$$\mathbb{P}\{M_n \leqslant -a \text{ et } S_n = b\} = \mathbb{P}\{S_n = -2a - b\}.$$

PREUVE. Par réflexion des chemins pour les temps $\geqslant T_{-a}$ au niveau $-a$, on obtient une bijection (principe de réflexion de D. André) entre

$$\{M_n \leqslant -a,\ S_n = b\} \quad \text{et} \quad \{S_n = -2a - b\}.$$



□

PROPOSITION 1.21. *Pour tout $a > 0$, on a*

$$\mathbb{P}\{T_{-a} \leqslant n\} = \mathbb{P}\big\{S_n \notin \,]-a, a]\big\}.$$

PREUVE. En effet, on a

$$\begin{aligned}
\mathbb{P}\{T_{-a} \leqslant n\} &= \mathbb{P}\{M_n \leqslant -a\} \\
&= \sum_{b=-\infty}^{\infty} \mathbb{P}\{M_n \leqslant -a,\ S_n = b\} \\
&= \sum_{b=-\infty}^{-a} \mathbb{P}\{S_n = b\} + \sum_{b=-a+1}^{\infty} \mathbb{P}\{S_n = -2a - b\} \quad \text{(par Corollaire 1.20)} \\
&= \mathbb{P}\{S_n \leqslant -a\} + \mathbb{P}\{S_n \leqslant -a - 1\} \\
&= \mathbb{P}\{S_n \leqslant -a\} + \mathbb{P}\{S_n \geq a + 1\} \quad \text{(par symétrie)} \\
&= \mathbb{P}\big\{S_n \notin \,]-a, a]\big\}.
\end{aligned}$$

□

COROLLAIRE 1.22. *Pour tout $a \neq 0$, on a $\mathbb{P}\{T_a < \infty\} = 1$, mais $\mathbb{E}[T_a] = \infty$.*

PREUVE. Sans restriction soit $a > 0$. Alors

$$\underbrace{\mathbb{P}\{T_{-a} > N\}}_{\downarrow\, \mathbb{P}\{T_{-a} = +\infty\}} = \mathbb{P}\big\{S_N \in \,]-a, a]\big\} \to 0, \quad N \to \infty.$$

Avec Exercise 1 du TD 7 (premier semestre) et Proposition 1.21 on obtient

$$\begin{aligned}
\mathbb{E}[T_a] &= \sum_{k=0}^{\infty} \mathbb{P}\{T_a > k\} = \sum_{k=0}^{\infty} \mathbb{P}\{T_{-a} > k\} \\
&= \sum_{k=0}^{\infty} \mathbb{P}\big\{S_k \in \,]-a, a]\big\} \\
&\geqslant \sum_{k=0}^{\infty} \mathbb{P}\{S_k = 0\} = +\infty,
\end{aligned}$$

car $\mathbb{P}\{S_{2k} = 0\} \sim \frac{1}{\sqrt{\pi k}}$. □

COROLLAIRE 1.23. *Pour $T_0 = \inf\{n > 0\colon\ S_n = 0\}$ on a:*

$$\mathbb{P}\{T_0 = \infty\} = 0 \quad et \quad \mathbb{E}[T_0] = +\infty.$$

PREUVE. En fait,

$$\begin{aligned}
\mathbb{P}\{T_0 \leqslant 2n\} &= \mathbb{P}\{T_0 \leqslant 2n;\ S_1 = +1\} + \mathbb{P}\{T_0 \leqslant 2n;\ S_1 = -1\} \\
&= \frac{1}{2}\mathbb{P}\{T_0 \leqslant 2n | S_1 = +1\} + \frac{1}{2}\mathbb{P}\{T_0 \leqslant 2n | S_1 = -1\} \\
&= \frac{1}{2}\mathbb{P}\{T_{-1} \leqslant 2n - 1\} + \frac{1}{2}\mathbb{P}\{T_1 \leqslant 2n - 1\} \\
&= \mathbb{P}\{T_{-1} \leqslant 2n - 1\},
\end{aligned}$$

et puis

$$\begin{aligned}
\mathbb{P}\{T_0 > 2n\} &= \mathbb{P}\{T_{-1} > 2n - 1\} \\
&= \mathbb{P}\{S_{2n-1} \in \,] - 1, 1]\} \quad \text{(par Proposition 1.21)} \\
&= \mathbb{P}\{S_{2n-1} = 1\} \\
&= \mathbb{P}\{S_{2n} = 0\} \to 0, \quad n \to \infty.
\end{aligned}$$

□

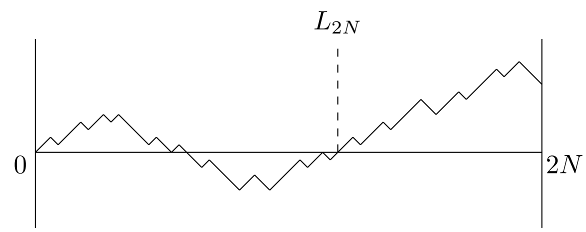
Interprétation. La marche aléatoire en dimension 1 est *récurrente*.

1.4. Loi discrète de l'arc-sinus pour la dernière visite

Considérons S_n, $n = 0, 1, \dots, 2N$. On note

$$L_{2N} = \max\bigl\{0 \leqslant n \leqslant 2N\colon \;\; S_n = 0\bigr\}$$

la "dernière visite en 0 avant $2N$" ("dernier retour à l'origine").



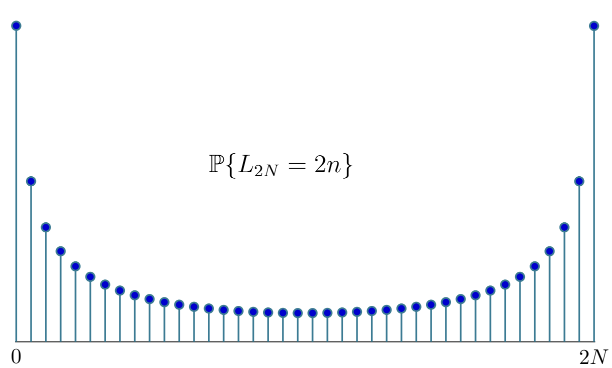
Interprétation Deux joueurs jouent à pile ou face 2N fois. Alors L_{2N} est la dernière fois qu'on a égalité entre les deux; après l'instant L_{2N} un de les deux est toujours avant l'autre.

Rappel On note $T_0 = \inf\{n > 0 : S_n = 0\}$. Pour tout n, on a

$$\mathbb{P}\{T_0 > 2n\} = \mathbb{P}\{S_{2n} = 0\}.$$

Pour $0 \leqslant n \leqslant N$ on obtient

$$\begin{aligned}
\mathbb{P}\{L_{2N} = 2n\} &= \mathbb{P}\bigl\{S_{2n} = 0,\;\; S_{2n+2} \neq 0, \dots, S_{2N} \neq 0\bigr\} \\
&= \mathbb{P}\bigl\{S_{2n+2} \neq 0, \dots, S_{2n} \neq 0 | S_{2n} = 0\bigr\}\mathbb{P}\{S_{2n} = 0\} \\
&= \mathbb{P}\{T_0 > 2N - 2n\}\,\mathbb{P}\{S_{2n} = 0\} \\
&= \mathbb{P}\{S_{2N-2n} = 0\}\,\mathbb{P}\{S_{2n} = 0\} \\
&= \binom{2N-2n}{N-n}\frac{1}{2^{2N-2n}}\binom{2n}{n}\frac{1}{2^{2n}} \\
&= \frac{\binom{2n}{n}\binom{2N-2n}{N-n}}{2^{2N}}.
\end{aligned}$$



PROPOSITION 1.24 (Loi discrète de l'arc-sinus pour la dernière visite).

$$\mathbb{P}\{L_{2n} = 2n\} = \mathbb{P}\{S_{2n} = 0\}\,\mathbb{P}\{S_{2(N-n)} = 0\}$$
$$= \frac{\binom{2n}{n}\binom{2N-2n}{N-n}}{2^{2N}}.$$

La loi de L_{2N} est symétrique par rapport à N, c.à.d.

$$\mathbb{P}\{L_{2N} = 2n\} = \mathbb{P}\{L_{2N} = 2(N-n)\}.$$

"Le gagnant prend la tête sans recours ou bien trés tôt ou bien très tard".

REMARQUE 1.25. *Pourquoi le nom arc-sinus*? On a

$$\mathbb{P}\{S_{2n} = 0\} \approx \frac{1}{\sqrt{\pi n}} \quad \text{et} \quad \mathbb{P}\{S_{2N-2n} = 0\} \approx \frac{1}{\sqrt{\pi(N-n)}},$$

et donc

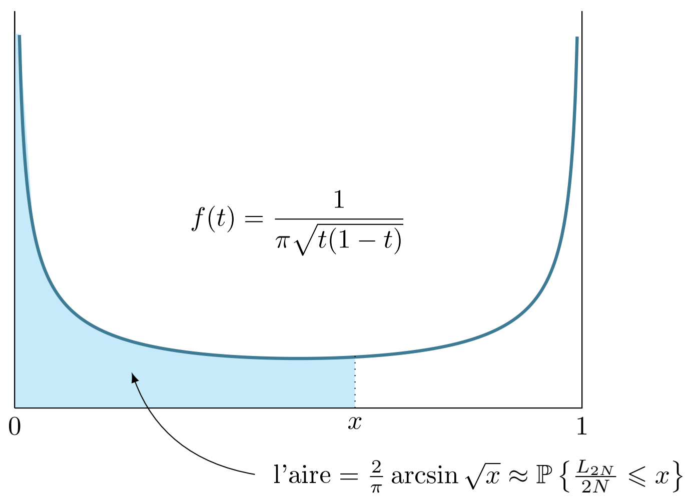
$$\mathbb{P}\{L_{2N} = 2n\} \approx \frac{1}{\pi\sqrt{n(N-n)}} = \frac{1}{N} f\left(\frac{n}{N}\right)$$

avec

$$f(t) = \frac{1}{\pi\sqrt{t(1-t)}}.$$

D'où on obtient pour $0 \leqslant x \leqslant 1$,

$$\mathbb{P}\left\{\frac{L_{2N}}{2N} \leqslant x\right\} = \sum_{n:\,\frac{n}{N}\leqslant x} \frac{1}{N} f\left(\frac{n}{N}\right) \approx \int_0^x f(t)dt = \frac{2}{\pi}\arcsin\sqrt{x}.$$



Interprétation. Si l'on fait un grand nombre de lancers à pile ou face, la dernière fois que le nombre de "pile" et le nombre de "face" obtenus ont coïncidé est proche du début ou de la fin de la série. Par exemple,

$$\mathbb{P}\{L_{10000} \leqslant 100\} \approx \frac{2}{\pi}\arcsin\sqrt{0.01} \approx 6.4\%$$
$$\mathbb{P}\{L_{10000} \leqslant 1000\} \approx \frac{2}{\pi}\arcsin\sqrt{0.1} \approx 20.5\%,$$

et par symétrie

$$\mathbb{P}\{L_{10000} \geqslant 9900\} \approx \frac{2}{\pi} \arcsin \sqrt{0.01} \approx 6.4\%$$
$$\mathbb{P}\{L_{10000} \geqslant 9000\} \approx \frac{2}{\pi} \arcsin \sqrt{0.1} \approx 20.5\%.$$

Remarque 1.26. Notons

$$\#\{\text{retours en } 0 \text{ de } S_n\} = \sum_{k=1}^{\infty} \mathbb{1}_{\{S_{2k}=0\}}.$$

Alors on a

$$\begin{aligned}
\mathbb{E}\big[\#\{\text{retours en } 0 \text{ de } S_n\}\big] &= \mathbb{E}\left[\sum_{k=1}^{\infty} \mathbb{1}_{\{S_{2k}=0\}}\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{1}_{\{S_{2k}=0\}}] \\
&= \sum_{k=1}^{\infty} \mathbb{P}\{S_{2k}=0\}.
\end{aligned}$$

Plus précisément,

$$\begin{aligned}
\tau_0 &:= 0 \\
\tau_1 &:= \inf\{k > 0 \colon\ S_k = 0\} \\
&\ \vdots \\
\tau_n &:= \inf\{k > \tau_{n-1} \colon\ S_k = 0\}.
\end{aligned}$$

Alors on obtient pour l'espérance de

$$V := \sum_{n \in \mathbb{N}} \mathbb{1}_{\{S_{2n}=0\}} = \sum_{n \in \mathbb{N}} \mathbb{1}_{\{\tau_n < \infty\}}$$

la formule

$$\mathbb{E}[V] = \sum_{n \in \mathbb{N}} \mathbb{P}_{\{\tau_n < \infty\}}.$$

D'autre part on a

$$\begin{aligned}
\mathbb{P}\{\tau_n < \infty\} &= \mathbb{P}\{\tau_{n-1} < \infty \quad \text{et} \quad \exists \ell : X_{\tau_{n-1}+1} + \cdots + X_{\tau_{n-1}+\ell} = 0\} \\
&= \mathbb{P}\{\tau_{n-1} < \infty\}\, \mathbb{P}\{\exists \ell : X_{\tau_{n-1}+1} + \cdots + X_{\tau_{n-1}+\ell} = 0\} \\
&= \mathbb{P}\{\tau_{n-1} < \infty\}\, \mathbb{P}\{\tau_1 < \infty\} \\
&= \mathbb{P}\{\tau_1 < \infty\}^n = 1.
\end{aligned}$$

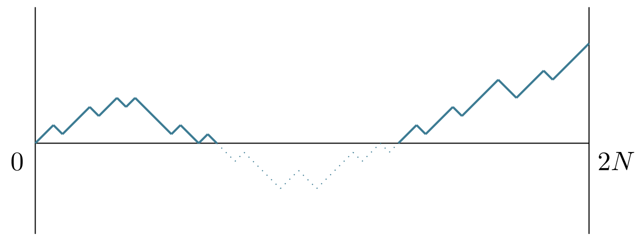
Par conséquent, on obtient que

$$\mathbb{P}\{S_n \text{ revient à l'origine infiniment souvent}\} = 1.$$

1.4.1. Loi discrète de l'arc-sinus pour les temps de séjour. On s'intérésse à la variable aléatoire

$$A_{2N}^{+} = \#\big\{1 \leqslant n \leqslant 2N\colon\ \max(S_{n-1}, S_n) > 0\big\},$$

"le temps total passé par la marche au-dessus de 0 pendant les $2N$ premiers pas".



Un chemin est dit *positif* [*resp. negatif*] sur l'interval $[n-1, n]$ si $S_{n-1}(\omega) > 0$ ou $S_n(\omega) > 0$ [resp. $S_{n-1}(\omega) < 0$ ou $S_n(\omega) < 0$]. On note

$$A_{2N,2n} = \big\{\text{chemins de longueur } 2N \underbrace{\text{avec } 2n \text{ trajets positifs}}_{(\text{et } 2(N-n) \text{ trajets négatifs})}\big\}.$$

Alors on a

$$\mathbb{P}\{A_{2N,2n}\} = \mathbb{P}\{A_{2N}^{+} = 2n\}, \quad n = 0, 1, \ldots, N.$$

Théorème 1.1 (Loi de l'arc-sinus pour les temps de séjour).
Pour $n = 0, 1, \ldots, N$, *on a*

$$(*) \qquad \mathbb{P}\{A_{2N}^{+} = 2n\} = \mathbb{P}\{S_{2n} = 0\}\,\mathbb{P}\{S_{2N-2n} = 0\}.$$

En particulier, pour tout $0 \leqslant \alpha < \beta \leqslant 1$,

$$\lim_{N\to\infty} \mathbb{P}\left\{\frac{A_{2N}^{+}}{2N} \in [\alpha, \beta]\right\} = \frac{2}{\pi}\left(\arcsin\sqrt{\beta} - \arcsin\sqrt{\alpha}\right).$$

Preuve. On remarque que par symétrie

$$\mathbb{P}\{A_{2N}^{+} = 2n\} = \mathbb{P}\{A_{2N}^{+} = 2N - 2n\}.$$

Soit $n = 0$ ou $n = N$:

$$\begin{aligned}
\mathbb{P}\{S_{2N} = 0\} &= \ \mathbb{P}\{T_0 > 2N\} \\
&= \ \mathbb{P}\{S_1 = 1, S_2 \geqslant 1, \ldots, S_{2N} \geqslant 1\} + \mathbb{P}\{S_1 = -1, S_2 \leqslant -1, \ldots, S_{2N} \leqslant -1\} \\
&= 2\mathbb{P}\{S_1 = 1, S_2 \geqslant 1, \ldots, S_{2N} \geqslant 1\} \\
&= 2\mathbb{P}\{S_2 \geqslant 1, \ldots, S_{2N} \geqslant 1 | S_1 = 1\}\mathbb{P}\{S_1 = 1\} \\
&= \ \mathbb{P}\{S_2 \geqslant 1, \ldots, S_{2N} \geqslant 1 | S_1 = 1\} \\
&= \ \mathbb{P}\{S_1 \geqslant 0, \ldots, S_{2N-1} \geqslant 0\} \\
&= \ \mathbb{P}\{S_1 \geqslant 0, \ldots, S_{2N-1} \geqslant 0, S_{2N} \geqslant 0\} \\
&= \ \mathbb{P}\{A_{2N}^{+} = 2N\} \\
&= \ \mathbb{P}\{A_{2N}^{+} = 0\}.
\end{aligned}$$

Il reste à démontrer $(*)$ pour tout $n \in \{1, \ldots, N-1\}$ et tout $N \geqslant 1$.

On procède avec *récurrence par rapport à* N (le cas $N = 1$ est déjà fait). Partition par rapport au premier retour à l'origine donne

$$\mathbb{P}\{A_{2n}^{+} = 2n\} = \mathbb{P}(A_{2N,2n}) = \sum_{r=1}^{N} \mathbb{P}\Big(A_{2N,2n} \cap \{T_0 = 2r\}\Big).$$

On decompose

$$\{T_0 = 2r\} = \{T_0 = 2r\}^{+} \cup \{T_0 = 2r\}^{-}$$

par rapport au sign sur $[0, 2r]$.

Alors

$$\begin{aligned}
\mathbb{P}\left(A_{2N,2n}\right) &= \sum_{r=1}^{N} \underbrace{\mathbb{P}\Big(A_{2N,2n} \cap \{T_0 = 2r\}^{+}\Big)}_{\neq 0 \text{ seulement pour } r \leqslant n} + \sum_{r=1}^{N} \underbrace{\mathbb{P}\Big(A_{2N,2n} \cap \{T_0 = 2r\}^{-}\Big)}_{\neq 0 \text{ seulement pour } r \leqslant N-n} \\
&= \sum_{r=1}^{N} \mathbb{P}\left(\{T_0 = 2r\}^{+}\right) \mathbb{P}\left(A_{2(N-r),2(n-r)}\right) + \sum_{r=1}^{N-n} \mathbb{P}\left(\{T_0 = 2r\}^{-}\right) \mathbb{P}\left(A_{2(N-r),2n}\right) \\
&= \frac{1}{2} \sum_{r=1}^{N} \mathbb{P}\{T_0 = 2r\}\, \mathbb{P}\left(A_{2(N-r),2(n-r)}\right) + \frac{1}{2} \sum_{r=1}^{N-n} \mathbb{P}\{T_0 = 2r\}\, \mathbb{P}\left(A_{2(N-r),2n}\right) \\
&= \frac{1}{2} \sum_{r=1}^{N} \mathbb{P}\{T_0 = 2r\}\, \mathbb{P}\left\{S_{2(n-r)=0}\right\} \mathbb{P}\left\{S_{2(N-n)=0}\right\} \\
&\quad + \frac{1}{2} \sum_{r=1}^{N-n} \mathbb{P}\{T_0 = 2r\}\, \mathbb{P}\left\{S_{2(N-r-n)=0}\right\} \mathbb{P}\left\{S_{2n=0}\right\} \quad \text{(par hypothèse de récurrence)} \\
&= \frac{1}{2} \mathbb{P}\{S_{2N-2n} = 0\} \underbrace{\sum_{r=1}^{n} \mathbb{P}\left\{T_0 = 2r\right\} \mathbb{P}\left\{S_{2n-2r} = 0\right\}}_{= \mathbb{P}\{S_{2n} = 0\}} \\
&\quad + \frac{1}{2} \mathbb{P}\{S_{2n} = 0\} \underbrace{\sum_{r=1}^{N-n} \mathbb{P}\left\{T_0 = 2r\right\} \mathbb{P}\left\{S_{2N-2n-2r} = 0\right\}}_{= \mathbb{P}\{S_{2N-2n} = 0\}} \\
&= \mathbb{P}\left\{S_{2n} = 0\right\} \mathbb{P}\left\{S_{2N-2n} = 0\right\}.
\end{aligned}$$

□

Interprétation On note

$$\frac{A_{2N}^{+}}{2N} \mathrel{\widehat{=}} \text{“la fraction du temps passé du coté positif”}.$$

Alors, pour N grand,

$$\mathbb{P}\left\{\frac{A_{2N}^{+}}{2N} \leqslant x\right\} \approx \frac{2}{\pi} \arcsin\sqrt{x}\,, \quad 0 \leqslant x \leqslant 1.$$

Par example, on obtient

$$\mathbb{P}\left\{\frac{A_{2N}^{+}}{2N} \leqslant 0.024\right\} \approx 0.1,$$

c.à.d.

$$\mathbb{P}\big\{\text{la marche passe au moins } 97.6\% \text{ de son temps du même côté de l'origine}\big\}$$
$$= \mathbb{P}\left(\left\{\frac{A_{2N}^{+}}{2N} \leqslant 0.024\right\} \cup \left\{\frac{A_{2N}^{+}}{2N} \geqslant 1 - 0.024\right\}\right) \approx 0.2$$

De même façon on obtient

$$\mathbb{P}\left\{\text{au moins } \underset{[99\%]}{85\%} \text{ de temps du même côté}\right\} \approx \underset{[0.13]}{0.5}$$

CHAPTER 2

Fonctions génératrices

Soit $a = (a_n)_{n \geqslant 0}$ une suite de nombres réels. On appelle *fonction génératrice* de la suite $a = (a_n)_{n \geqslant 0}$ la fonction

$$G_a(s) = \sum_{n=0}^{\infty} a_n s^n$$

pour $s \in \mathbb{C}$ tels que la série converge.

- *Convergence.* Il existe $0 \leqslant R \leqslant \infty$ (*rayon de convergence*) tel que la série converge absolument si $|s| < R$, et diverge si $|s| > R$.
- *Différentiation.* Sur $|s| < R$, G_a peut être différentiée terme à terme.
- *Unicité.* S'il existe $r > 0$ tel que $G_a(s) = G_b(s)$ pour $s \in [-r, r]$, alors $a_n = b_n$ pour tout n. De plus *(formule de Cauchy)*,
$$a_n = \frac{G_a^{(n)}}{n!}(0).$$
- *Continuité (Abel).* Si $a_n \geqslant 0$ pour tout n, et $G_a(s)$ converge pour $|s| < 1$, alors
$$\lim_{\mathbb{R} \ni s \uparrow 1} G_a(s) = \sum_{n=0}^{\infty} a_n$$
que cette somme soit finie ou égale à $+\infty$.

Définition 2.1. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On appelle *fonction génératrice* de X la fonction

$$G_X(s) := \mathbb{E}[s^X] = \sum_{n=0}^{\infty} \underbrace{\mathbb{P}\{X = n\}}_{=p_n} s^n, \quad s \in \mathbb{C}.$$

Remarque 2.2. Le rayon de convergence R de G_X est toujours $\geqslant 1$. En effet, pour $-1 \leqslant s \leqslant 1$, on a

$$\big|G_X(s)\big| \leqslant \mathbb{E}\big[\,|s|^X\,\big] \leqslant \mathbb{E}\,[1] = 1.$$

De plus,

$$G_X(1-) := \lim_{\mathbb{R} \ni s \uparrow 1} G_X(s) = \sum_{n=0}^{\infty} p_n = 1.$$

Remarque 2.3. La fonction génératrice G_X (et même sa restriction à l'intervalle $[-1, 1]$) caractérise la loi de X. En effet,

$$\mathbb{P}\{X = n\} = \frac{G_X^{(n)}}{n!}(0).$$

EXEMPLE 2.4. (1) Soit $X \sim \mathcal{B}(1,p)$ *(Bernoulli de paramètre p)*, c.à.d.

$$\mathbb{P}\{X=1\} = 1 - \mathbb{P}\{X=0\} = p.$$

Alors on a

$$G_X(s) = s^0 \underbrace{\mathbb{P}\{X=0\}}_{=1-p=q} + s^1 \underbrace{\mathbb{P}\{X=1\}}_{=p} = q + ps.$$

(2) Soit $X \sim \mathcal{B}(n,p)$ *(loi binomiale de paramètres n et p)*. Alors on a

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{n} \mathbb{P}\{X=k\} s^k = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k \\ &= ((1-p)+ps)^n = (q+ps)^n. \end{aligned}$$

(3) Soit $X \sim \mathcal{P}(\lambda)$ *(loi de Poisson de paramètre $\lambda > 0$)*. Alors on a

$$G_X(s) = \sum_{n=0}^{\infty} \mathbb{P}\{X=n\} s^n = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} s^n = e^{\lambda(s-1)}.$$

(4) Soit $X \sim \mathcal{G}(p)$ *(loi géométrique de paramètre p)*. Alors on a

$$\begin{aligned} G_X(s) &= \sum_{n=0}^{\infty} \mathbb{P}\{X=n\} s^n = \sum_{n=1}^{\infty} p(1-p)^{n-1} s^n \\ &= ps \sum_{n=1}^{\infty} (1-p)^{n-1} s^{n-1} \\ &= ps \sum_{n=0}^{\infty} ((1-p)s)^n = \frac{ps}{1-(1-p)s} \end{aligned}$$

pour tout s tel que si $(1-p)|s| < 1$, i.e. $|s| < \frac{1}{1-p}$.

REMARQUE 2.5. *(Fonction génératrice et moments)* Soit G_X la fonction génératrice d'une variable aléatoire X à valeurs dans $\mathbb{N}$, c.à.d.

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n, \quad |s| < 1,$$

avec $p_n = \mathbb{P}\{X=n\}$. Alors on a pour $|s| < 1$,

$$\begin{aligned} G'_X(s) &= \sum_{n=0}^{\infty} n p_n s^{n-1}, \\ G''_X(s) &= \sum_{n=0}^{\infty} n(n-1) p_n s^{n-2}, \\ &\vdots \\ G^{(\ell)}_X(s) &= \sum_{n=0}^{\infty} n \cdot \ldots \cdot (n-\ell+1) p_n s^{n-\ell}. \end{aligned}$$

En particulier, on obtient

$$G'_X(1-) = \sum_{n=0}^{\infty} np_n = \mathbb{E}[X], \quad si \quad \mathbb{E}[X] < \infty,$$

$$G''_X(1-) = \sum_{n=0}^{\infty} n(n-1)p_n = \mathbb{E}[X(X-1)], \quad si \quad \mathbb{E}[X(X-1)] < \infty,$$

$$\dots$$

$$G^{(\ell)}_X(1-) = \sum_{n=0}^{\infty} n \cdot \ldots \cdot (n-\ell+1)p_n = \mathbb{E}[X(X-1)\cdot \ldots \cdot (X-\ell+1)],$$

$$\text{si } \mathbb{E}[X(X-1)\cdot \ldots \cdot (X-\ell+1)] < \infty.$$

Proposition 2.6. *On a*

$$\mathbb{E}[X] = G'_X(1-)$$
$$\mathbb{E}[X^2] = G''_X(1-) + G'_X(1-),$$

où les membres de gauche sont finis si et seulement si les membres de droite sont finis. En particuliers, si $G'_X(1-) < \infty$ et $G''_X(1-) < \infty$, on a pour la variance:

$$\mathrm{var}(X) = G''_X(1-) + G'_X(1-) - G'_X(1-)^2.$$

Exemple 2.7.

(1) Soit $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$. Alors on a

$$G_X(s) = e^{\lambda(s-1)}, \quad s \in \mathbb{C}.$$

Par conséquent, on obtient

$$G'_X(s) = \lambda e^{\lambda(s-1)}, \ldots, G^{(\ell)}_X(s) = \lambda^{\ell} e^{\lambda(s-1)},$$

et d'où

$$\mathbb{E}[X] = G'_X(1) = \lambda,$$
$$\vdots$$
$$\mathbb{E}[X(X-1)\cdot \ldots \cdot (X-\ell+1)] = \lambda^{\ell}.$$

En particulier, on a

$$\mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - E[X] = \lambda^2,$$

et donc

$$\mathrm{var}(X) = \underbrace{\mathbb{E}[X^2]}_{=\lambda^2+\lambda} - \underbrace{\mathbb{E}[X]^2}_{=\lambda^2} = \lambda.$$

(2) Soit $X \sim \mathcal{G}(p)$, $0 < p < 1$. Alors on a

$$G_X(s) = \frac{p^s}{1-qs}, \quad |s| < \frac{1}{q}, \quad \text{avec } q = 1-p.$$

On trouve

$$G'_X(s) = \frac{p(1-qs) + pqs}{(1-qs)^2} = \frac{p}{(1-qs)^2},$$
$$G''_X(s) = \frac{2pq}{(1-qs)^3}.$$

Pour $s = 1$ il resulte

$$G'_X(1) = \mathbb{E}[X] = \frac{1}{p}$$
$$G''_X(1) = \mathbb{E}[X^2] - E[X] = \frac{2q}{p^2}.$$

Par conséquent, on obtient

$$\mathbb{E}[X^2] = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{2-p}{p^2},$$

et finalement,

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

PROPOSITION 2.8. *Si X et Y sont deux variables aléatoires indépendantes, alors*

$$G_{X+Y} = G_X G_Y.$$

PREUVE. En utilisant l'indépendance de X et Y, on a

$$G_{X+Y}(s) = \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] = \mathbb{E}[s^X]\,\mathbb{E}[s^Y] = G_X(s)\,G_Y(s). \qquad \square$$

EXEMPLE 2.9. Soit

$$X = X_1 + \ldots + X_n,$$

avec (X_i) i.i.d. et $\mathbb{P}\{X_i = 1\} = 1 - \mathbb{P}\{X_i = 0\} = p$. Alors

$$X \sim \mathcal{B}(n,p),$$

et on obtient

$$G_X(s) = G_{X_1}(s) \cdot \ldots \cdot G_{X_n}(s) = G_{X_1}(s)^n = (q+ps)^n, \quad s \in \mathbb{C}.$$

EXEMPLE 2.10. *(Loi binomiale négative)* Soit $p \in]0,1[$. On regarde un jeu de pile ou face:

$$\begin{cases} \textit{pile} \ \text{ avec probabilité } \ p, \\ \textit{face} \ \text{ avec probabilité } \ 1-p = q. \end{cases}$$

On note Z_r le nombre de jets nécessaires pour obtenir r fois pile, $r \geqslant 1$. En posant

$$\begin{cases} X_1 = Z_1 \\ X_2 = Z_2 - Z_1 \\ \quad \vdots \\ X_r = Z_r - Z_{r-1} \\ \quad \vdots \end{cases}$$

on obtient une suite (X_i) i.i.d. telle que $X_i \sim \mathcal{G}(p)$. En particulier,

$$G_{X_i}(s) = \frac{ps}{1-qs}, \quad |s| < \frac{1}{q}.$$

Puisque $Z_r = X_1 + \ldots + X_r$, on trouve

$$G_{Z_r}(s) = G_{X_1}(s) \cdot \ldots \cdot G_{X_r}(s) = G_{X_1}(s)^r = \left(\frac{ps}{1-qs}\right)^r.$$

Rappel Pour $\alpha \in \mathbb{R}$ et $|x| < 1$ on a

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k$$

où les coefficients binomiaux $C_\alpha^k \equiv \binom{\alpha}{k}$ sont définies par

$$\binom{\alpha}{0} := 1 \quad \text{et}$$
$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)\cdot\ldots\cdot(\alpha-k+1)}{k!}.$$

On note que

$$C_{-r}^k = \binom{-r}{k} = (-1)^k \binom{r+k-1}{r-1} = (-1)^k C_{r+k-1}^{r-1}, \quad 1 \leqslant r \in \mathbb{N}.$$

En effet,

$$\begin{aligned}\binom{-r}{k} &= \frac{(-r)(-r-1)\cdot\ldots\cdot(-r-k+1)}{k!} \\ &= (-1)^k \frac{(r+k-1)\cdot\ldots\cdot(r+1)r}{k!} \\ &= (-1)^k \frac{(r+k-1)!}{(r-1)!\,k!} \\ &= (-1)^k \binom{r+k-1}{r-1} = (-1)^k C_{r+k-1}^{r-1}.\end{aligned}$$

En termes de ce développement, on obtient

$$(1-qs)^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} (-qs)^k = \sum_{k=0}^{\infty} C_{r+k-1}^{r-1} q^k s^k$$

et

$$G_{Z_r}(s) = \sum_{k=0}^{\infty} C_{r+k-1}^{r-1} p^r q^k s^{r+k}.$$

D'où on déduit que

$$G_{Z_r}^{(n)}(s) = \sum_{k=0}^{\infty} C_{r+k-1}^{r-1} p^r q^k \, n! \, s^{r+k-n} \, \mathbb{1}_{\{k \geq n-r\}},$$

et donc

$$\mathbb{P}\{Z_r = n\} = \frac{G_{Z_r}^{(n)}(0)}{n!} = \begin{cases} 0 & \text{pour } n < r \\ C_{n-1}^{r-1}\, p^r q^{n-r} & \text{pour } n \geqslant r \end{cases}$$

(loi binômiale négative).

2.1. Théorème de continuité

Théorème 2.11 (Théorème de continuité). *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires à valeur dans $\mathbb{N}$ et X aussi une variable aléatoire à valeurs dans $\mathbb{N}$. Les deux conditions suivantes sont équivalentes:*

(1) *Pour tout $k \in \mathbb{N}$ on a $\mathbb{P}\{X_n = k\} \longrightarrow \mathbb{P}\{X = k\}$, quand $n \to \infty$*
(2) *Pour tout $s \in \,]0,1[$ on a $G_{X_n}(s) \longrightarrow G_X(s)$, quand $n \to \infty$.*

PREUVE. (1) $\Rightarrow$ (2): On sait que

$$\underbrace{\mathbb{P}\{X_n = k\}}_{=:p_k^n} \longrightarrow \underbrace{\mathbb{P}\{X = k\}}_{=:p_k}, \quad n \to \infty$$

On fixe $\varepsilon > 0$. Choisissons $N \in \mathbb{N}$ tel que

$$\sum_{k=N}^{\infty} p_k \leqslant \frac{\varepsilon}{4}$$

et puis n suffisament grand tel que

$$|p_k^n - p_k| \leqslant \frac{\varepsilon}{4N} \quad \text{pour tout } k = 0, 1, \ldots, N-1.$$

Alors

$$\sum_{k=N}^{\infty} p_k^n = 1 - \sum_{k=0}^{N-1} p_k^n \leqslant 1 - \sum_{k=0}^{N-1} p_k + \frac{\varepsilon}{4} = \sum_{k=N}^{\infty} p_k + \frac{\varepsilon}{4} \leqslant \frac{\varepsilon}{2}.$$

Donc, pour n suffisamment grand et pour $s\colon |s| \leqslant 1$, on obtient

$$\begin{aligned}
|G_{X_n}(s) - G_X(s)| &\leqslant \sum_{k=0}^{\infty} |p_k^n - p_k||s|^k \\
&\leqslant \sum_{k=0}^{\infty} |p_k^n - p_k| \\
&\leqslant \underbrace{\sum_{k=0}^{N-1} |p_k^n - p_k|}_{\leqslant \varepsilon/4} + \underbrace{\sum_{k=N}^{\infty} p_k^n}_{\leqslant \varepsilon/2} + \underbrace{\sum_{k=N}^{\infty} p_k}_{\leqslant \varepsilon/4} \leqslant \varepsilon.
\end{aligned}$$

(2) $\Rightarrow$ (1): Pour $0 < s < 1$ on a

$$G_{X_n}(s) \to G_X(s), \quad n \to \infty,$$

c.à.d.

$$\sum_{k=0}^{\infty} p_k^n s^k \to \sum_{k=0}^{\infty} p_k s^k, \quad n \to \infty.$$

D'abord nous démontrons que

$$p_0^n \to p_0, \quad \text{quand } n \to \infty,$$

c.à.d.

$$P\{X_n = 0\} \to \mathbb{P}\{X = 0\}, \quad \text{quand } n \to \infty.$$

En effet, on observe que $G'_{X_n} \geqslant 0$ sur $]0,1[$ et donc $G_{X_n}\!\uparrow$ sur $]0,1[$. Pour tout $0 < s < 1$ on obtient

$$\mathbb{P}\{X_n = 0\} = p_0^n = G_{X_n}(0) \leqslant G_{X_n}(s) = \sum_{k \geqslant 0} p_n^k s^k \leqslant p_0^n + \sum_{k \geqslant 1} s^k = p_0^n + \frac{s}{1-s},$$

et donc

$$G_{X_n}(s) - \frac{s}{1-s} \leqslant \mathbb{P}\{X_n = 0\} \leqslant G_{X_n}(s).$$

On déduit que

$$G_X(s) - \frac{s}{1-s} \leqslant \liminf_{n\to\infty} \mathbb{P}\{X_n = 0\} \leqslant \limsup_{n\to\infty} \mathbb{P}\{X_n = 0\} \leqslant G_X(s)$$

et puis, en passant à la limite $s \downarrow 0$,

$$\lim_{n\to\infty} \mathbb{P}\{X_n = 0\} = G_X(0) = \mathbb{P}\{X = 0\}.$$

On procède maintenant par récurrence. Supposons que $p_k^n \to p_k$ pour $k < r$. Alors, on a pour $0 < s < 1$,

$$\frac{G_{X_n}(s) - \sum_{k=1}^{r-1} p_k^n s^k}{s^r} \quad \stackrel{n\to\infty}{\longrightarrow} \quad \frac{G_X(s) - \sum_{k=1}^{r-1} p_k s^k}{s^r},$$

c.à.d.

$$H^n(s) := \sum_{k\geqslant 0} p_{k+r}^n s^k \quad \stackrel{n\to\infty}{\longrightarrow} \quad \sum_{k\geqslant 0} p_{k+r} s^k =: H(s).$$

On peut donc répéter le raisonnement précédent pour obtenir:

$$H^n(0) = p_r^n \quad \stackrel{n\to\infty}{\longrightarrow} \quad H(0) = p_r. \qquad \square$$

EXEMPLE 2.12. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires telle que $X_n \sim \mathcal{B}(n, p_n)$ et

$$\lim_{n\to\infty} np_n = \lambda > 0.$$

Alors

$$\begin{aligned}\lim_{n\to\infty} G_{X_n}(s) &= \lim_{n\to\infty} (1 + (s-1)p_n)^n \\ &= \lim_{n\to\infty} \left(1 + \frac{(s-1)\,np_n}{n}\right)^n \\ &= e^{(s-1)\lambda} = G_X(s)\end{aligned}$$

pour une variable aléatoire $X \sim \mathcal{P}(\lambda)$. Par le Théorème de continuité, on a donc

$$\mathbb{P}\{X_n = k\} \stackrel{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}.$$

2.2. Théorème limite (approximation poissonienne)

THÉORÈME 2.13 (Approximation poissonienne). *Soit*

$$(X_{n,i})_{n\in\mathbb{N}^*,\, i\in\mathbb{N}^*,\, 1\leqslant i\leqslant i_n}$$

un "schéma en triangle" ($i_1 \leqslant i_2 \leqslant i_3 \leqslant \ldots$) de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$ tel que

(i) *chaque $X_{n,i}$ suit une loi de Bernoulli:*

$$p_{n,i} := \mathbb{P}\{X_{n,i} = 1\} = 1 - \mathbb{P}\{X_{n,i} = 0\};$$

(ii) *$X_{n,1}, X_{n,2}, \ldots, X_{n,i_n}$ sont indépendantes pour tout n;*
(iii) *il existe $\lambda > 0$ tel que*

$$p_{n,1} + \ldots + p_{n,i_n} \longrightarrow \lambda, \quad \textit{quand } n \to \infty$$

$$\textit{et} \quad \max_{1\leqslant i\leqslant n} p_{n,i} \longrightarrow 0, \quad \textit{quand } n \to \infty.$$

Alors pour $S_n = X_{n,1} + \ldots + X_{n,i_n}$ et tout $k \geqslant 0$ on a:

$$\mathbb{P}\{S_n = k\} \longrightarrow e^{-\lambda}\frac{\lambda^k}{k!}, \quad \textit{quand } n \to \infty$$

PREUVE. Il faut démontrer que $G_{S_n}(s) \to e^{\lambda(s-1)}$ pour tout $s \in]0,1[$. En effet, on a

$$G_{X_{n,i}}(s) = \mathbb{E}[s^{X_{n,i}}] = p_{n,i}s + (1-p_{n,i}),$$

et donc

$$G_{S_n}(s) = \prod_{i=1}^{i_n} G_{X_{n,i}}(s) = \prod_{i=1}^{i_n}(p_{n,i}s + (1-p_{n,i})).$$

D'où on obtient

$$\log G_{S_n}(s) = \sum_{i=1}^{i_n} \log(p_{n,i}s + (1-p_{n,i})) = \sum_{i=1}^{i_n} \log(1 - p_{n,i}(s-1)).$$

Utilisant le fait que par Taylor $\log(1-x) = -x + \theta(x)x$ avec $\theta(x) \to 0$, quand $x \to 0$, et comme $\theta(p_{n,i}(1-s)) \leqslant \varepsilon$ pour $n \geqslant n(\varepsilon)$ uniformément en i et s, on obtient

$$\log G_{S_n}(s) = \underbrace{-\sum_{i=1}^{i_n} p_{n,i}(1-s)}_{\to \lambda(1-s)} + \underbrace{\sum_{i=1}^{i_n} \theta(p_{n,i}(1-s))p_{n,i}(1-s).}_{\leqslant \varepsilon(1-s)\sum_{i=1}^{i_n} p_{n,i} \to \varepsilon(1-s)\lambda}$$

Ceci montre que

$$\log G_{S_n}(s) \to \lambda(s-1), \quad n \to \infty,$$

et donc

$$G_{S_n}(s) \to e^{\lambda(s-1)}. \quad n \to \infty.$$

□

EXEMPLE 2.14 (Cas particulier). Soit $p_{n,i} = p_n$ tel que $np_n \to \lambda$, et $i_n := n$. Alors $S_n \sim \mathcal{B}(n, p_n)$, et on retrouve que

$$\mathbb{P}\{S_n = k\} \to e^{-\lambda}\frac{\lambda^k}{k!}, \quad n \to \infty.$$

PROPOSITION 2.15 (Somme aléatoire des variables aléatoires). *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de même loi, à valeurs dans $\mathbb{N}$, et soit T une variable aléatoire aussi à valeurs dans $\mathbb{N}$. On suppose que toutes les variables $T, X_1, X_2, \ldots$ sont indépendantes. On pose*

$$S_n := \sum_{k=1}^{n} X_k \quad \big(\textit{c.à.d.} \quad S_n(\omega) = \sum_{k=1}^{n} X_k(\omega),\ \omega \in \Omega\big), \ \textit{et}$$
$$S_T := \sum_{k=1}^{T} X_k \quad \big(\textit{c.à.d.} \quad S_T(\omega) = \sum_{k=1}^{T(\omega)} X_k(\omega),\ \omega \in \Omega\big).$$

(On utilise la convention que $\sum_{k=1}^{T(\omega)} X_k(\omega) := 0$ pour $T(\omega) = 0$). Alors

$$G_{S_T} = G_T \circ G_{X_1}.$$

PREUVE. Pour $|s| < 1$ on a

$$\begin{aligned} G_{S_T}(s) &= \sum_{k=0}^{\infty} \mathbb{P}\{S_T = k\} s^k \\ &= \sum_{k=0}^{\infty} s^k \sum_{n=0}^{\infty} \mathbb{P}\{S_T = k,\, T = n\} \end{aligned}$$

La famille

$$\left(s^k\, \mathbb{P}\{S_T = k, T = n\}\right)_{(k,n)\in\mathbb{N}\times\mathbb{N}}$$

est absolument sommable (uniformement en s): en effet

$$\sum_{k,n} |s|^k\, \mathbb{P}\{S_T = k, T = n\} \leqslant \sum_{k,n} \mathbb{P}\{S_T = k,\, T = n\} = 1.$$

Si on utilise en plus que $\mathbb{P}\{S_T = k,\, T = n\} = \mathbb{P}\{S_T = k\}\mathbb{P}\{T = n\}$, on peut donc calculer

$$\begin{aligned} G_{S_T}(s) =& \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} s^k \mathbb{P}\{S_n = k, T = n\} \right) \\ =& \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} s^k \mathbb{P}\{S_n = k\} \right) \mathbb{P}\{T = n\} \\ =& \sum_{n=0}^{\infty} G_{S_n}(s)\, \mathbb{P}\{T = n\} \\ =& \sum_{n=0}^{\infty} G_{X_1}(s)^n\, \mathbb{P}\{T = n\} \\ =& \mathbb{E}\left[G_{X_1}(s)^T\right] = G_T(G_{X_1}(s)). \end{aligned}$$

□

COROLLAIRE 2.16 (Identité de Wald). *Supposons les mêmes hypothèses que dans la Proposition.*

(1) *Si* $\mathbb{E}[X_1] < \infty$ *et* $\mathbb{E}[T] < \infty$, *alors* $\mathbb{E}[S_T] < \infty$ *et* $\mathbb{E}[S_T] = \mathbb{E}[X_1]\,\mathbb{E}[T]$.
(2) *Si* X_1 *et* T *admettent un moment d'ordre deux, il en est de même pour* S_T, *et on a:*

$$\operatorname{var}(S_T) = \operatorname{var}(X_1)\mathbb{E}[T] + \mathbb{E}[X_1]^2 \operatorname{var}(T).$$

PREUVE.

(1) On sait que

$$G_{S_T}(s) = G_T(G_{X_1}(s)), \quad |s| < 1.$$

Par differentiation on obtient

$$G'_{S_T}(s) = G'_T(G_{X_1}(s)) \cdot G'_{X_1}(s), \quad -1 < s < 1.$$

Utilisant que $G_{X_1}(s) \uparrow 1$ quand $s \uparrow 1$, on déduit que

$$\underbrace{G'_{S_T}(1-)}_{=\mathbb{E}[S_T]} = \underbrace{G'_T(1-)}_{=\mathbb{E}[T]} \cdot \underbrace{G'_{X_1}(1-)}_{=\mathbb{E}[X_1]}.$$

(2) Par differentiation de la relation $G'_{S_T}(s) = G'_T(G_{X_1}(s)) \cdot G'_{X_1}(s)$ une autre fois, on obtient

$$G''_{S_T}(s) = G''_T(G_{X_1}(s)) \cdot G'_{X_1}(s)^2 + G'_T(G_{X_1}(s)) \cdot G''_{X_1}(s)^2, \quad s \in\,]-1, 1[.$$

Pour $s \uparrow 1$, on en déduit que

$$G''_{S_T}(1-) = G''_T(1-) \cdot G'_{X_1}(1-)^2 + G'_T(1-) \cdot G''_{X_1}(1-)^2.$$

Autrement dit, on a

$$\mathbb{E}[S_T(S_T-1)] = \mathbb{E}[T(T-1)]\,\mathbb{E}[X_1]^2 + \mathbb{E}[T]\,\mathbb{E}[X_1(X_1-1)].$$

Avec ceci, on obtient

$$\begin{aligned}
\mathrm{var}(S_T) &= \mathbb{E}[S_T(S_T-1)] + \mathbb{E}[S_T] - \mathbb{E}[S_T]^2 \\
&= (\mathbb{E}[T^2] - \mathbb{E}[T])\,\mathbb{E}[X_1]^2 + \mathbb{E}[T]\,(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]) + \mathbb{E}[X_1]\mathbb{E}[T] - \mathbb{E}[X_1]^2\mathbb{E}[T]^2 \\
&= (\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2)\,\mathbb{E}[T] + \mathbb{E}[X_1]^2\,(\mathbb{E}[T^2] - \mathbb{E}[T]^2) \\
&= \mathrm{var}(X_1)\,\mathbb{E}[T] + \mathbb{E}[X_1]^2\ \mathrm{var}(T).
\end{aligned}$$

□

2.3. Processus de branchement: le processus de Galton-Watson

Le modèle de Galton-Watson a été introduit en 1873 dans une correspondance entre Sir Francis Galton et Révérend Henry Williams Watson en vue d'étudier la statistique des noms de famille dans l'Angleterre victorienne et en particulier à l'extinction de ces noms. On a cherché une explication de la disparition progressive des grands noms de l'aristocratie anglaise.

Hypothèses:

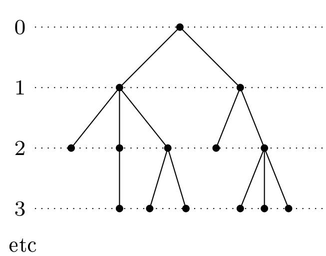
(1) Une population évolue par générations: on note Z_n le nombre d'individus de la $n^{\text{ième}}$ génération.
(2) Chaque membre de la $n^{\text{ième}}$ génération donne naissance à une "famille" des descendantes mâles (éventuellement vide) de la generation suivante:

$$\begin{aligned}
Z_0 &= 1 \\
Z_{n+1} &= X_1^{n+1} + \ldots + X_{Z_n}^{n+1}, \quad n \geq 0.
\end{aligned}$$

Les tailles X_k^n de chaque famille forment des variables aléatoires indépendantes; toutes de même loi. On note

$$\mu := \mathbb{E}[X_k^n] \text{ et } \sigma^2 = \mathrm{var}(X_k^n).$$

Par example,



$$Z_0(\omega) = 1$$

$$Z_1(\omega) = X_1^1(\omega) = 2$$

$$Z_2(\omega) = X_1^2(\omega) + X_2^2(\omega) = 3 + 2 = 5$$

$$Z_3(\omega) = \underbrace{X_1^3(\omega)}_{=0} + \underbrace{X_2^3(\omega)}_{=1} + \underbrace{X_3^3(\omega)}_{=2} + \underbrace{X_4^3(\omega)}_{=0} + \underbrace{X_5^3(\omega)}_{=3} = 6.$$

REMARQUE 2.17. Utilisant la notion

$$G_n(s) := G_{Z_n}(s) \quad \text{et} \quad G(s) := G_{X_k^n}(s),$$

on obtient

$$\begin{aligned} &G_1(s) = G(s), \quad \text{et} \\ &G_n(s) = (G_{n-1} \circ G)(s) = \underbrace{(G \circ \ldots \circ G)}_{\text{n-fois}}(s)\,, \quad n \geqslant 2. \end{aligned}$$

En plus, on observe

$$\mathbb{E}[Z_n] = \underbrace{\mathbb{E}[X_1^n]}_{=\mu} \mathbb{E}[Z_{n-1}] = \ldots = \mu^n$$

et

$$\operatorname{var}(Z_n) = \underbrace{\operatorname{var}(X_1^n)}_{=\sigma^2} \underbrace{\mathbb{E}[Z_{n-1}]}_{=\mu^{n-1}} + \underbrace{\mathbb{E}[X_1^n]}_{=\mu^2} \operatorname{var}(Z_{n-1}) = \sigma^2 \mu^{n-1} + \mu^2 \operatorname{var}(Z_{n-1}).$$

COROLLAIRE 2.18. *On a*

$$\operatorname{var}(Z_n) = \begin{cases} n\sigma^2 & \textit{si} \quad \mu = 1, \\ \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \textit{si} \quad \mu \neq 1. \end{cases}$$

PREUVE. 1. Soit $\mu = 1$. Alors on trouve

$$\operatorname{var}(Z_n) = \sigma^2 + \operatorname{var}(Z_{n-1}) = \ldots = n\sigma^2 + \underbrace{\operatorname{var}(Z_0)}_{=0}.$$

2. Soit maintenant $\mu \neq 1$. On procède par récurrence.

$n = 0:\quad \operatorname{var}(Z_0) = 0.$

$n - 1 \mapsto n\ :$

$$\begin{aligned} \operatorname{var}(Z_n) &= \sigma^2 \mu^{n-1} + \mu^2 \operatorname{var}(Z_{n-1}) \\ &= \sigma^2 \mu^{n-1} + \frac{\sigma^2(\mu^{n-1} - 1)\mu^n}{\mu - 1} \\ &= \sigma^2\, \frac{\mu^n - \mu^{n-1} + \mu^{2n-1} - \mu^n}{\mu - 1} \\ &= \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1}. \end{aligned}$$

□

Problème. On remarque que

$$\begin{aligned} \{Z_n = 0\} \uparrow \bigcup_{n=1}^{\infty} \{Z_n = 0\} &= \{Z_n = 0 \text{ pour un } n \in \mathbb{N}^*\} \\ &= \left\{\lim_{n\to\infty} Z_n = 0\right\} = \{Z_n \to 0\} \quad \text{“\textit{extinction}”}. \end{aligned}$$

On a donc

$$\mathbb{P}\{Z_n = 0\} \uparrow \mathbb{P}\{Z_n \to 0\}.$$

Question : Comment calculer la probabilité

$$\eta := \mathbb{P}\{Z_n \to 0\} = \mathbb{P}\{\text{“extinction”}\}?$$

THÉORÈME 2.1. *Soit* $\mu = \mathbb{E}[Z_1] < \infty$ *et* $0 < \sigma^2 = \mathrm{var}(Z_1) < \infty$.

(1) *La probabilité d'extinction*

$$\eta = \mathbb{P}\{\exists n \in \mathbb{N}\colon Z_n = 0\} = \mathbb{P}\left\{\lim_{n\to\infty} Z_n = 0\right\}$$

est donnée par la plus petite racine positive de l'équation

$$s = G(s).$$

(2) *En plus, on a*

$$\begin{aligned} \mu \leqslant 1 &\Rightarrow \eta = 1 \\ \mu > 1 &\Rightarrow \eta < 1. \end{aligned}$$

PREUVE. (1) On remarque que la suite

$$\eta_n = \mathbb{P}\{Z_n = 0\}$$

est croissante et bornée par 1. Donc η_n est une suite convergente. De plus,

$$\eta_n \uparrow \mathbb{P}\left(\cup_n \{Z_n = 0\}\right) = \eta.$$

Puisque G est continue, on obtient

$$\underbrace{\eta_n}_{\to\,\eta} = G_n(0) = G(G_{n-1}(0)) = \underbrace{G(\eta_{n-1})}_{\to\,G(\eta)}.$$

Par conséquent, η est une racine positive de l'équation $s = G(s)$.

(2) If faut démontrer que η est la plus petite racine positive de l'équation

$$s = G(s).$$

Supposons que α soit une autre racine positive de $s = G(s)$. Il faut montrer que $\eta \leqslant \alpha$. En effet, comme

$$G'(s) = \mathbb{E}[Z_1 s^{Z_1}] = \sum_{k\geqslant 1} k s^{k-1} \mathbb{P}\{Z_1 = k\} \geqslant 0, \quad \text{si } s \geqslant 0,$$

on note que $G|[0,1]$ est une fonction croissante. Donc on a

$$\begin{aligned} \eta_1 &= G(0) \ \leqslant G(\alpha) = \alpha \\ \eta_2 &= G(\eta_1) \leqslant G(\alpha) = \alpha \\ &\vdots \quad \text{etc} \end{aligned}$$

Par récurrence, on obtient

$$\eta_n \leqslant \alpha \quad \text{pour tout } n,$$

et donc

$$\eta \leqslant \alpha.$$

Autrement dit, η est bien la plus petite racine positive de l'équation $s = G(s)$.

(3) On remarque en plus que la fonction G est convexe sur $[0,1]$:

$$\begin{aligned} G''(s) &= \mathbb{E}[Z_1(Z_1 - 1)s^{Z_1 - 2}] \\ &= \sum_{k\geqslant 2}^{\infty} k(k-1)s^{k-2}\mathbb{P}\{Z_1 = k\} \geqslant 0, \quad \text{si } s \geqslant 0. \end{aligned}$$

Si $\mu \geqslant 1$, la fonction G est même strictement convexe, car

$$\mathbb{P}\{Z_1 \geqslant 2\} > 0.$$

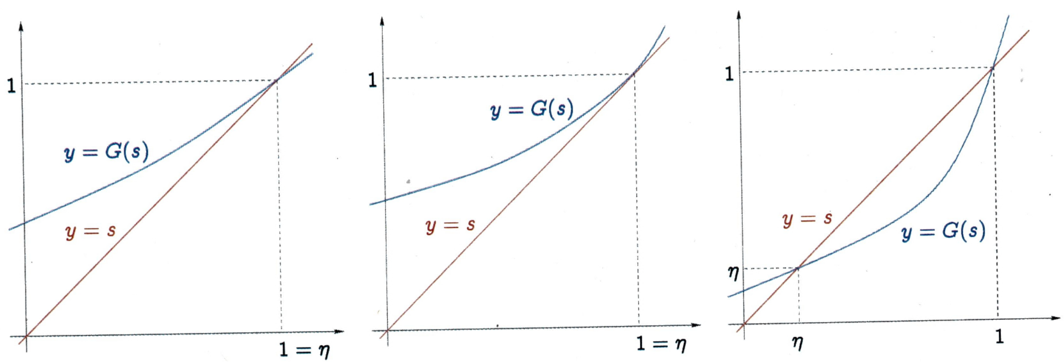
(En effet, supposons que $\mathbb{P}\{Z_1 \geqslant 2\} = 0$, alors $\mathbb{P}\{Z_1 = 1\} = 1$ car $\mathbb{E}[Z_1] \geqslant 1$, et donc $\sigma^2 = \mathrm{var}(Z_1) = 0$). De plus on a

$$G(1) = 1 \quad \text{et} \quad G'(1-) = \mu.$$

Les deux courbes $y = G(s)$ et y=s ont généralement deux points d'intersection:

$$\eta \;\text{ et }\; 1.$$

Rappellons que $G(0) = \mathbb{P}\{Z_1 = 0\}$.



$\mu < 1$ ("subcritical") $\mu = 1$ ("critical") $\mu > 1$ ("supercritical")

$\boxed{\mu < 1}$: $G'(1-) = \mu < 1$; les deux points d'intersection η et 1 coïncident;

$\boxed{\mu = 1}$: G est strictement convexe et $G'(1-) = 1$, donc $G(s) > s$ pour $0 \leqslant s < 1$, et $\eta = 1$;

$\boxed{\mu > 1}$: on a $0 \leqslant \eta < 1$, et $\eta = 0$ si et seulement si $\mathbb{P}\{Z_1 = 0\} = 0$. □

Exemple 2.19. *(Probabilité de survie des familles nobles)* Le nom de famille est transmis à tout les enfants mâles par leur père; le nombre de fils d'une famille est une variable aléatoire à valeurs dans $\mathbb{N}$. On pose

$$p_\ell := \mathbb{P}\{Z_1 = \ell\} = \mathbb{P}\{X_1^1 = \ell\}, \quad \ell = 0, 1, 2, \ldots$$

Supposons que

$$p_0 = \frac{1}{8}, \quad p_1 = \frac{3}{8}, \quad p_2 = \frac{3}{8}, \quad p_3 = \frac{1}{8}.$$

Alors

$$\begin{aligned} \mathbb{E}[Z_1] = \mathbb{E}[X_1^1] &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} \\ &= \frac{12}{8} = \frac{3}{2} = \mu > 1 \end{aligned}$$

et

$$\begin{aligned}G(s) = \mathbb{E}[s^{Z_1}] &= p_0 + p_1 s + p_2 s^2 + p_3 s^3\\ &= \frac{1}{8} + \frac{3}{8}s + \frac{3}{8}s^2 + \frac{1}{8}s^3.\end{aligned}$$

Donc, on obtient:

$$\begin{aligned}s \text{ point fixe de } G &\iff s = G(s)\\ &\iff s = \frac{1}{8} + \frac{3}{8}s + \frac{3}{8}s^2 + \frac{1}{8}s^3\\ &\iff 1 - 5s + 3s^2 + s^3 = 0\\ &\iff (s-1)(s^2 + 4s - 1) = 0.\end{aligned}$$

Trivialement, $s = 1$ est a point fixe; les deux autres solutions sont $s = -2 \pm \sqrt{5}$. Par conséquent,

$$\eta = -2 + \sqrt{5}.$$

Conclusion. On trouve

$$\mathbb{P}\{\text{“\textit{extinction du nom de famille}”}\} = -2 + \sqrt{5} \approx 0.27$$

CHAPTER 3

Fonctions génératrices des moments et le théorème de Cramér

3.1. Propriétés des fonctions génératrices des moments

DÉFINITION 3.1. La *fonction génératrice des moments* d'une variable aléatoire réelle X est la fonction

$$M_X : \mathbb{R} \to]0, \infty], \quad M_X(t) = \mathbb{E}[e^{tX}].$$

On remarque que pour une variable aléatoire X à valeurs dans $\mathbb{N}$, on a

$$M_X(t) = G_X(e^t).$$

En effet, pour tout t tel que $e^t <$ rayon de convergence de G_X, on a

$$G_X(e^t) = \mathbb{E}[(e^t)^X] = \mathbb{E}[e^{tX}] = M_X(t).$$

REMARQUE 3.2 (Propriétés élémentaires).

(i) On a toujours $M_X(0) = 1$.
(ii) Si X est bornée, alors M_X est définie et continue sur $\mathbb{R}$.
(iii) Si la loi de X est symétrique (c.à.d. si X et $-X$ ont même loi), alors M_X est une fonction paire.
(iv) Pour tout $a, b \in \mathbb{R}$,

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

PROPOSITION 3.3. *Supposons qu'il existe $t_0 > 0$ tel que*

$$M_X(t) < \infty \quad \textit{pour tout } t \in]-t_0, t_0[.$$

(1) *Alors $\mathbb{E}[|X|^n] < \infty$ pour tout $n \in \mathbb{N}$.*
(2) *On a*

$$M_X(t) = \sum_{n=0}^{\infty} \mathbb{E}[X^n] \frac{t^n}{n!}, \quad |t| < t_0.$$

En particulier,

$$M_X^{(n)}(0) = \mathbb{E}[X^n].$$

PREUVE.

(1) Utilisant que pour $0 < t < t_0$,

$$|x|^n \leqslant \frac{n!}{t^n} e^{t|x|} \leqslant \frac{n!}{t^n} \left(e^{t\,x} + e^{-t\,x} \right), \quad n \geqslant 1,$$

on a

$$\mathbb{E}\left[|X|^n\right] \leqslant \frac{n!}{t^n} \big(M_X(t) + M_X(-t) \big), \quad 0 < t < t_0.$$

(2) On observe que pour $|t| < t_0$:

$$\sum_{n=0}^{N} \mathbb{E}[X^n] \frac{t^n}{n!} = \mathbb{E}\left[\sum_{n=0}^{N} \frac{(tX)^n}{n!}\right] \xrightarrow{N\to\infty} \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] = \mathbb{E}[e^{tX}].$$

Pour la convergence on utilise la convergence dominée:

$$\sum_{n\in\mathbb{N}} \frac{|tX|^n}{n!} = e^{|tX|} \leqslant e^{tX} + e^{-tX}.$$ □

PROPOSITION 3.4. *Soient X et Y deux variables aléatoires indépendantes. Alors*

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

PREUVE. On a

$$\mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} e^{tY}] = \mathbb{E}[e^{tX}]\, \mathbb{E}[e^{tY}].$$ □

REMARQUE 3.5. Si X est une variable aléatoire continue avec densité f_X, alors

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\; dx$$

La fonction

$$L_X(t) := M_X(-t) = \int_{-\infty}^{\infty} e^{-tx} f_X(x)\; dx$$

s'appelle *transformée de Laplace* de f_X.

Résultat d'analyse. La fonction f_X est uniquement déterminée par sa transformée de Laplace L_X. Plus généralement, on peut démontrer la proposition suivante.

PROPOSITION 3.6 (Théorème d'unicité). *Soit $t_0 > 0$. Soient X et Y deux variables aléatoires telles que*

$$M_X(t) = M_Y(t) < \infty, \quad |t| < t_0.$$

Alors X et Y ont la même loi.

EXEMPLE 3.7. a) Soit X une variable normale $N(0,1)$. Alors

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

et on obtient

$$M_X(t) = \mathbb{E}[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2+tx}\, dx.$$

Utilisant que

$$-\frac{x^2}{2} + tx = -\frac{1}{2}(x-t)^2 + \frac{t^2}{2},$$

on déduit

$$\begin{aligned} M_X(t) &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx \\ &= e^{t^2/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\, du}_{=1} = e^{t^2/2}, \quad \text{avec } u = x - t. \end{aligned}$$

b) Dans le cas général d'une variable X normale $N(\mu, \sigma^2)$, on peut écrire

$$X = \sigma G + \mu \quad \text{avec } G \text{ une variable } N(0,1).$$

D'où on déduit

$$M_X(t) = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}, \quad t \in \mathbb{R}.$$

COROLLAIRE 3.8. *Pour une variable X qui suit la loi $N(\mu, \sigma^2)$ on a*

$$\begin{aligned} M_X'(t) &= (\mu + \sigma^2 t)\, M_X(t), \quad \textit{et} \\ M_X''(t) &= \big(\sigma^2 + (\mu + \sigma^2 t)^2\big)\, M_X(t), \end{aligned}$$

en particulier

$$\begin{aligned} M_X'(0) &= \mu = \mathbb{E}[X] \\ M_X''(0) &= \sigma^2 + \mu^2 = \mathbb{E}[X^2]. \end{aligned}$$

D'où on retrouve bien la variance d'une variable aléatoire suivant la loi $N(\mu, \sigma^2)$:

$$\operatorname{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

3.2. Inégalités de Chernoff

DÉFINITION 3.9. Soit X une variable aléatoire réelle telle que

$$M_X(t) = \mathbb{E}[e^{tX}] < \infty \quad \text{pour tout } t \in \mathbb{R}.$$

On appelle transformée de Cramér de X la fonction

$$I\colon\ \mathbb{R} \to [0, \infty], \quad I_X(a) = \begin{cases} \sup\limits_{t \geqslant 0} \big(at - \log M_X(t)\big), & a \geqslant 0, \\ \sup\limits_{t \leqslant 0} \big(at - \log M_X(t)\big), & a \leqslant 0. \end{cases}$$

On appelle $\Lambda_X(t) := \log M_X(t)$ la transformée log-Laplace de X.

THÉORÈME 3.1. *(Inégalités de Chernoff) Soit X une variable aléatoire telle que*

$$M_X(t) < \infty \quad \textit{pour } t \in \mathbb{R}.$$

Alors on a, pour tout $a \geqslant 0$,

$$\begin{cases} \mathbb{P}\{X \geqslant a\} \leqslant e^{-I_X(a)} \\ \mathbb{P}\{X \leqslant -a\} \leqslant e^{-I_X(-a)}. \end{cases}$$

PREUVE. Fixons $a \geqslant 0$. Soit $t \geqslant 0$. Alors

$$e^{t(X-a)} \geqslant \mathbb{1}_{\{X \geqslant a\}}$$

et donc

$$\mathbb{E}[e^{t(X-a)}] \geqslant \mathbb{P}\{X \geqslant a\}.$$

Utilisant que

$$\mathbb{E}[e^{t(X-a)}] = e^{-at}\, \mathbb{E}[e^{tX}] = e^{-(at - \Lambda_X(t))} \quad \text{avec } \Lambda_X(t) = \log M_X(t),$$

on obtient

$$\mathbb{P}\{X \geqslant a\} \leqslant e^{-(at - \Lambda_X(t))}.$$

et donc

$$\mathbb{P}\{X \geqslant a\} \leqslant \exp\left(-\sup_{t \geqslant 0} \big(at - \Lambda_X(t)\big)\right) = e^{-I_X(a)}.$$

Soit maintenant $t \leqslant 0$. De la manière analogue, on obtient

$$\mathbb{P}\{X \leqslant -a\} = \mathbb{E}[\mathbb{1}_{\{X \leqslant -a\}}] \leqslant \mathbb{E}[e^{t(X+a)}] = e^{ta+\Lambda_X(t)} = e^{-(-at-\Lambda_X(t))},$$

et d'où

$$\mathbb{P}\{X \leqslant -a\} \leqslant \exp\Big(-\sup_{t \leqslant 0}\big(-at - \Lambda_X(t)\big)\Big) = e^{-I_X(-a)}. \qquad \square$$

Proposition 3.10. *Soit* $(X_n)_{n \geqslant 1}$ *une suite de variables aléatoires i.i.d. de même loi que* X *et supposons que*

$$M_X(t) < \infty \quad \textit{pour tout } t \in \mathbb{R}.$$

Notons

$$S_n := X_1 + \ldots + X_n$$

et

$$\bar{X} := \frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}$$

la moyenne empirique. Alors on a

$$I_{\bar{X}}(a) = n\, I_X(a), \quad a \in \mathbb{R}.$$

Preuve. On a

$$\begin{aligned} M_{\bar{X}t} &= \mathbb{E}\left[e^{t\frac{X_1+\ldots+X_n}{n}}\right] = \mathbb{E}\left[e^{\frac{t}{n}X_1+\ldots+\frac{t}{n}X_n}\right] \\ &= \prod_{i=1}^{n} \mathbb{E}\left[e^{\frac{t}{n}X_i}\right] = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \\ &= M_X\left(\frac{t}{n}\right)^n. \end{aligned}$$

On obtient

$$\Lambda_{\bar{X}(t)} = n\, \Lambda_X\left(\frac{t}{n}\right),$$

et puis pour $a \geqslant 0$,

$$\begin{aligned} I_{\bar{X}}(a) &= \sup_{t \geqslant 0}\left(at - n\Lambda_X\left(\frac{t}{n}\right)\right) \\ &= n \sup_{t \geqslant 0}\left(a\frac{t}{n} - \Lambda_X\left(\frac{t}{n}\right)\right) \\ &= n\, I_X(a). \end{aligned}$$

Pour $a \leqslant 0$, on a la même relation avec la même démonstration. $\square$

Corollaire 3.11 (Théorème de Cramér, 1938). *Sous les conditions de Proposition* 3.10 *on a, avec* $I = I_{X_1}$,

$$\left.\begin{aligned} \mathbb{P}\left\{\frac{S_n}{n} \geqslant a\right\} &\leqslant e^{-nI(a)} \\ \mathbb{P}\left\{\frac{S_n}{n} \leqslant -a\right\} &\leqslant e^{-nI(-a)} \end{aligned}\right\} \quad \textit{pour tout } a \geqslant 0.$$

LEMME 3.12. *Considérons la fonction*

$$\rho\colon t \longmapsto ta - \log \mathbb{E}[e^{tX}] = ta - \log M_X(t).$$

On a les propriétés suivantes:

(1) $\rho(0) = 0$;

(2) $\rho'(t) = a - \dfrac{M_X'(t)}{M_X(t)} = a - \dfrac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}$; *en particulier,* $\rho'(0) = a - \mathbb{E}[X]$.

(3) *La fonction* ρ *est concave,*

$$\begin{aligned}\rho''(t) &= \frac{d}{dt}\left(a - \frac{M_X'(t)}{M_X(t)}\right) = -\left(\frac{M_X''(t)}{M_X(t)} - \frac{M_X'(t)^2}{M_X(t)^2}\right)\\ &= -\underbrace{\left(\frac{\mathbb{E}[X^2e^{tX}]}{\mathbb{E}[e^{tX}]} - \left(\frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}\right)^2\right)}_{\geqslant 0} \leqslant 0.\end{aligned}$$

En effet, par Cauchy-Schwarz

$$\big(\mathbb{E}[Xe^{tX}]\big)^2 = (\mathbb{E}[e^{X/2}\cdot Xe^{X/2}])^2 \leqslant \mathbb{E}[e^{tX}]\mathbb{E}[X^2e^{tX}].$$

EXEMPLE 3.13. Soit $a > \mathbb{E}[X]$. Alors $\rho(0) = 0$, $\rho'(0) > 0$ et ρ est concave.

DÉFINITION 3.14. Soit X une variable aléatoire réelle telle que

$$M_X(t) = \mathbb{E}[e^{tX}] < \infty \quad \text{pour tout } t \in \mathbb{R}.$$

On pose

$$\Lambda_X(t) = \log(M_X(t)) \quad \text{(transformée log-Laplace)}$$

et

$$\hat{I}_X(x) = \sup_{t\in\mathbb{R}} \big(tx - \Lambda_X(t)\big), \quad x \in \mathbb{R} \quad \text{(transformée de Legendre de } \Lambda_X).$$

COROLLAIRE 3.15 (Théorème de Cramér-Chernoff). *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires i.i.d. telle que* $M_{X_1}(t) < \infty$ *pour tout* $t \in \mathbb{R}$. *On note*

$$\mu = \mathbb{E}[X_1], \quad \hat{I} = \hat{I}_{X_1} \quad \textit{et} \quad \Lambda = \Lambda_{X_1}.$$

Soit $S_n = X_1 + \ldots + X_n$. *Alors on a*

(i) $\mathbb{P}\left\{\dfrac{S_n}{n} \geqslant x\right\} \leqslant e^{-n\hat{I}(x)}$ *pour tout* $x \in [\mu, \infty[$;

(ii) $\mathbb{P}\left\{\dfrac{S_n}{n} \leqslant x\right\} \leqslant e^{-n\hat{I}(x)}$ *pour tout* $x \in\,]-\infty, \mu]$.

PREUVE. Sans restrictions, on peut se ramener au cas $\mu = 0$, sinon on passe de X_n à $X_n^* := X_n - \mathbb{E}[X_n] = X_n - \mu$. En effet, si par example $x \geqslant \mu$, alors avec $S_n^* = X_1^* + \ldots + X_n^*$,

$$\mathbb{P}\left\{\frac{S_n}{n} \geqslant x\right\} = \mathbb{P}\left\{\frac{S_n}{n} - \mu \geqslant x - \mu\right\} = \mathbb{P}\left\{\frac{S_n^*}{n} \geqslant x - \mu\right\} \overset{!}{\leqslant} e^{-n\hat{I}_{X_1^*}(x-\mu)},$$

mais

$$\begin{aligned}\hat{I}_{X_1^*}(x-\mu) &= \sup_{t\in\mathbb{R}} \big(tx - t\mu - \log(\mathbb{E}[e^{-t\mu}e^{tX_1}])\big)\\ &= \sup_{t\in\mathbb{R}} (tx - t\mu + t\mu - \Lambda(t))\\ &= \hat{I}(x).\end{aligned}$$

Donc soit maintenant $x \in [\mu, \infty[$ et sans restrictions $\mu = 0$. Par le Théorème de Cramér, on a

$$\mathbb{P}\left\{\frac{S_n}{n} \geqslant x\right\} \leqslant e^{-nI(x)}$$

où $I(x) = \sup_{t \geqslant 0}(tx - \Lambda_{X_1}(t))$, car $x \geqslant 0$. La fonction

$$\rho\colon t \mapsto tx - \log(\Lambda_{X_1}(t))$$

a les propriétés

$$\begin{cases} \rho(0) = 0 \\ \rho'(0) = x - \mu = x \geqslant 0 \\ \rho \text{ est concave,} \end{cases}$$

d'où on déduit

$$I(x) = \sup_{t \geqslant 0} \rho(t) = \sup_{t \in \mathbb{R}} \rho(t) = \hat{I}(x).$$

De même façon, on montre (ii). $\square$

Corollaire 3.16 (Loi forte des grands nombres). *Soit $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires i.i.d. telles que $M_{X_1}(t) < \infty$ pour tout $t \in \mathbb{R}$. Alors*

$$\frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n} \longrightarrow \mathbb{E}[X_1] \quad p.s.$$

Preuve. Sans restrictions $\mathbb{E}[X_1] = 0$. De grandes déviations à la loi des grands nombres: Pour tout $\varepsilon > 0$, on a

$$\mathbb{P}\left\{\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right\} \leqslant e^{-n\hat{I}(\varepsilon)} + e^{-n\hat{I}(-\varepsilon)}.$$

Puisque $\hat{I}(\varepsilon)$ et $\hat{I}(-\varepsilon)$ sont strictement positifs, on déduit

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right\} < \infty \quad \text{(convergence rapide en probabilité).}$$

Avec le Lemme de Borel-Cantelli on obtient la conclusion:

$$\frac{S_n}{n} \to 0 \quad \text{p.s.} \tag{3.1}$$

$\square$

CHAPTER 4

Vecteurs gaussiens

Un *vecteur aléatoire* à n dimensions est une variable aléatoire X à valeurs dans $\mathbb{R}^n$. On ecrit

$$X = (X_1, \ldots, X_n).$$

On fait l'hypothèse que

$$\boxed{\mathbb{E}[X_i^2] < \infty \quad \forall i = 1, \ldots, n.}$$

Alors on a $\mathbb{E}\,|X_i X_j| < \infty$ pour tout i, j.

4.1. Matrices de variance-covariance

La matrice

$$\operatorname{var}(X) = (\operatorname{cov}(X_i, X_j))_{1 \leqslant i,j \leqslant n}$$

s'appelle *matrice de variance-covariance* de X.

On a

$$\operatorname{var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^t (X - \mathbb{E}[X])\big] = \mathbb{E}\big[X^t X\big] - \mathbb{E}[X]^t \mathbb{E}[X]$$

où

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots \mathbb{E}[X_n]).$$

Remarque 4.1 (Transformation linéaire). Soit $A \in \operatorname{Matr}(m \times n; \mathbb{R})$ et $b \in \mathbb{R}^m$. Alors le vecteur aléatoire à m dimensions

$$Y = AX + b$$

a pour espérance

$$\mathbb{E}[Y] = A\,\mathbb{E}[X] + b$$

et pour matrice de variance-covariance

$$\operatorname{var}(Y) = A \operatorname{var}(\mathrm{X}) A^t.$$

En effet,

$$\begin{aligned} \operatorname{var}(Y)_{ij} = \operatorname{cov}(Y_i, Y_j) &= \operatorname{cov}\left(\sum_k A_{ik} X_k, \sum_\ell A_{j\ell} X_\ell\right) \\ &= \sum_{k,\ell} A_{ik} \operatorname{cov}(X_k, X_\ell) A^t_{\ell j} \\ &= \left(A \operatorname{var}(\mathrm{X}) A^t\right)_{ij}. \end{aligned}$$

REMARQUE 4.2. Soient X et Y deux vecteurs aléatoires à valeurs dans $\mathbb{R}^n$ et $\mathbb{R}^m$ respectivement,

$$X = (X_1, \dots, X_n) \text{ et } Y = (Y_1, \dots, Y_m).$$

On définit la *matrice de covariance* comme

$$\text{cov(X,Y)} = (\text{cov}(X_i, Y_j))_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant m}} \in \text{Matr}(n \times m; \mathbb{R}).$$

REMARQUE 4.3. Si $Z = (X_1, \dots, X_n, Y_1, \dots, Y_m)$, on a

$$\text{var}(Z) = \begin{array}{cc} & \begin{array}{cc} n & m \end{array} \\ \begin{array}{c} n \\ m \end{array} & \begin{pmatrix} \text{var}(X) & \text{cov}(X,Y) \\ \text{cov}(Y,X) & \text{var}(Y) \end{pmatrix} \end{array} \in \text{Matr}((n+m) \times (n+m); \mathbb{R}).$$

On remarque que $\text{cov}(Y, X) = \text{cov}(X, Y)^t$.

DÉFINITION 4.4 (vecteur gaussien). Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de $\mathbb{R}^n$ est dit *gaussien* si pour tout $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, la variable aléatoire

$$\langle a, X \rangle = \sum_{i=1}^{n} a_i X_i$$

est une variable réelle gaussienne.

REMARQUE 4.5. Si $X = (X_1, \dots, X_n)$ est gaussien à valeurs dans $\mathbb{R}^n$, alors pour tout $1 \leqslant p \leqslant n$ et tout $(i_1, \dots, i_p) \in \{1, \dots, n\}^p$, le vecteur aléatoire $(X_{i_1}, \dots, X_{i_p})$ est gaussien.

DÉFINITION 4.6. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. On appelle fonction génératrice des moments de X la fonction

$$M_X(a) := \mathbb{E}[e^{\langle a, X \rangle}], \quad a \in \mathbb{R}^n.$$

REMARQUE 4.7. Soit M_X la fonction génératrice des moments d'un vecteur aléatoire X. Si $M_X < \infty$, alors M_X caractérise la loi de X. En plus,

$$(X_1, \dots, X_n) \text{ indépendent } \iff M_X(a) = M_{X_1}(a_1) \cdot \ldots \cdot M_{X_n}(a_n) \quad \forall a \in \mathbb{R}^n.$$

PROPOSITION 4.8. *Soit $X = (X_1, \dots, X_n)$ un vecteur gaussien avec $\mu = \mathbb{E}[X]$ la moyenne et $C = \text{var}(X)$ la matrice de variance-covariance. Alors*

$$M_X(a) = e^{\langle a, \mu \rangle + \frac{1}{2} \langle a, Ca \rangle}, \quad a \in \mathbb{R}^n.$$

PREUVE. On utilise que

$$\langle a, X \rangle \sim N(\langle a, \mu \rangle, \langle a, Ca \rangle).$$

Alors, pour $t \in \mathbb{R}$ on a

$$M_X(ta) = \mathbb{E}[e^{t\langle a, X \rangle}] = M_{\langle a, X \rangle}(t) = \exp\Big(\langle a, \mu \rangle t + \frac{1}{2} \langle a, Ca \rangle t^2\Big).$$

Il suffit de prendre $t = 1$. □

COROLLAIRE 4.9. *Deux vecteurs gaussiens*

$$X = (X_1, \dots, X_n) \quad \textit{et} \quad Y = (Y_1, \dots, Y_n)$$

ont même loi si et seulement si ils ont même vecteur moyenne et même matrice de covariance, c.à.d. si et seulement si $\mathbb{E}[X_i] = \mathbb{E}[Y_i]$ et $\text{cov}(X_i, X_j) = \text{cov}(Y_i, Y_j)$ pour tout i, j. En d'autres termes, la loi d'un vecteur gaussien est caracterisée par $\mu = \mathbb{E}[X]$ et $C = \text{var}(X)$.

PROPOSITION 4.10 (vecteurs gaussiens et indépendance). *Soit* $X = (X_1, \ldots, X_n)$ *un vecteur gaussien. Ses composantes* $X_1, \ldots, X_n$ *sont indépendantes si et seulement si* $X_1, \ldots, X_n$ *sont non-correlées, c.à.d. si et seulement si la matrice de covariance* $\mathrm{var}(X)$ *est diagonale.*

PREUVE. Soit $\mu = \mathbb{E}[X]$ et $C = \mathrm{var}(X)$. Si $\mathrm{var}(X)$ est diagonale, alors

$$\langle a, Ca \rangle = \sum_i C_{ii} a_i^2 = \sum_i \mathrm{var}(X_i) a_i^2.$$

Par conséquent,

$$\begin{aligned} M_X(a) &= e^{\langle a,\mu\rangle + \frac{1}{2}\langle a, Ca\rangle} \\ &= e^{\sum\limits_i a_i \mathbb{E}[X_i] + \frac{1}{2}\sum\limits_i a_i^2 \,\mathrm{var}(X_i)} \\ &= \prod_{i=1}^n e^{a_i \mathbb{E}[X_i] + \frac{1}{2} a_i^2 \,\mathrm{var}(X_i)} \\ &= \prod_{i=1}^n M_{X_i}(a_i) = M_{X_1}(a_1) \cdot \ldots \cdot M_{X_n}(a_n). \end{aligned}$$

Donc $(X_1, \ldots, X_n)$ est indépendant. $\square$

PROPOSITION 4.11 (existence d'un vecteur gaussien). *Soient* $\mu \in \mathbb{R}^n$ *et* $C \in \mathrm{Matr}(n \times n;\ \mathbb{R})$. *Alors, il existe un espace de probabilité* $(\Omega, \mathscr{A}, \mathbb{P})$ *et un vecteur gaussien défini sur* $(\Omega, \mathscr{A}, \mathbb{P})$ *de vecteur moyenne* μ *et de matrice de variance-covariance* C *si et seulement si* C *est symétrique positive.*

PREUVE. "$\Rightarrow$" Soit $X = (X_1, \ldots, X_n)$ gaussien et $C = \mathrm{var}(X)$. Alors $C_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{cov}(X_j, X_i) = C_{ji}$ et $\langle a, Ca \rangle = \mathrm{var}(\langle a, X \rangle) \geqslant 0\ \forall a \in \mathbb{R}^n$.
"$\Leftarrow$" Soit $C \in \mathrm{Matr}(n \times n; \mathbb{R})$ symétrique. Alors:

$$\begin{aligned} C \text{ positive } &\Leftrightarrow \exists A \in \mathrm{Matr}(n \times n;\ \mathbb{R}) \text{ telle que } C = A \cdot A^t \\ &\Leftrightarrow \langle a, Ca \rangle \geqslant 0, \quad \forall a \in \mathbb{R}^n. \end{aligned}$$

Soient $X_1, \ldots, X_n$ indépendantes telles que $X_i \sim N(0,1)$ pour $i = 1, \ldots, n$. Considérons $X = (X_1, \ldots X_n)$. Alors

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]) = 0$$

et

$$\mathrm{var}(X) = \mathbb{1}_n = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Donc,

$$Y := A \cdot X + \mu$$

a pour espérance

$$\mathbb{E}[Y] = A \cdot \underbrace{\mathbb{E}[X]}_{=0} + \mu = \mu$$

et pour matrice variance-covariance

$$\mathrm{var}(Y) = A \underbrace{\mathrm{var}(X)}_{=\mathbb{1}_n} A^t = A \cdot A^t = C.$$ $\square$

Notation 4.12. Pour une matrice $C \in \mathrm{Matr}(n \times n; \mathbb{R})$ symétrique positive et un vecteur $\mu \in \mathbb{R}^n$ on note

- $N(\mu, C)$ la *loi gaussienne* sur $\mathbb{R}^n$ de *moyenne* μ et de *matrix de variance-covariance* C;
- $N(0, \mathbb{1}_n)$ la *loi gaussienne standard* sur $\mathbb{R}^n$.

4.2. Vecteur gaussien et densité

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire à n dimensions. La loi de X est dite *absolument continue*, si'l existe une densité $f_X \colon \mathbb{R}^n \to \mathbb{R}_+$ mesurable telle que

$$\begin{aligned}\mathbb{E}[h(X)] &= \int_{\mathbb{R}^n} h(x) f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(x_1, \dots, x_n) f_X(x_1, \dots, x_n)\, dx_1 \dots dx_n\end{aligned}$$

pour toute fonction $h \colon \mathbb{R}^n \to \mathbb{R}_+$ mesurable.

Par exemple, pour

$$h = \mathbb{1}_{]-\infty, a_1]} \times \dots \times \mathbb{1}_{]-\infty, a_n]},$$

on obtient

$$\begin{aligned}\mathbb{E}[h(X)] &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f_X(x_1, \dots x_n)\, dx_1 \dots dx_n \\ &= \mathbb{P}\{X_1 \leqslant a_1, \dots, X_n \leqslant a_n\}.\end{aligned}$$

Dans ce cas $\mathbb{E}[h(X)]$ est la fonction de répartition de X:

$$F_X(a) = \mathbb{P}\{X_1 \leqslant a_1, \dots, X_n \leqslant a_n\}, \quad a = (a_1, \dots, a_n) \in \mathbb{R}^n.$$

Proposition 4.13. *Soit $X = (X_1, \dots, X_n)$ gaussien de moyenne μ et de matrice variance-covariance C. Alors:*

(1) *La loi de X est absolument continue si et seulement si $C = \mathrm{var}(X)$ inversible.*
(2) *Si C est inversible, alors X admet pour densité la fonction*

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \exp\left(-\frac{\langle x - \mu, C^{-1}(x - \mu)\rangle}{2}\right), \quad x \in \mathbb{R}^n.$$

Preuve. Si la matrice C est symétrique positive, alors il existe une matrice $A \in \mathrm{Matr}(n \times n; \mathbb{R})$ telle que

$$C = A \cdot A^t.$$

Soit $Z = (Z_1, \dots, Z_n) \sim N(0, \mathbb{1}_n)$. Alors X a même loi que $AZ + \mu$. Les $Z_1, \dots, Z_n$ sont des variables aléatoires réelles i.i.d. de loi $N(0,1)$. Donc $Z = (Z_1, \dots, Z_n)$ a pour densité la fonction

$$f_Z : \mathbb{R}^n \to \mathbb{R}_+, \quad f_Z(x) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2/2}, \quad x \in \mathbb{R}^n.$$

Considérons maintenant la fonction

$$g \colon \mathbb{R}^n \to R^n, \quad z \mapsto Az + \mu.$$

- Supposons que C soit inversible, alors A est aussi inversible, car

$$0 \neq \det C = (\det A)^2.$$

Donc $g\colon \mathbb{R}^n \to \mathbb{R}^n$ est une bijection, C^∞, de matrice Jacobienne

$$\mathcal{J}_g(z) = A.$$

Donc,

$$\begin{aligned}
\mathbb{E}[h(X)] &= \mathbb{E}[h \circ g(Z)] \\
&= \int_{\mathbb{R}^n} h(g(x)) f_Z(x)\, dx \\
&= \int_{\mathbb{R}^n} h(x) f_Z(g^{-1}(x)) \frac{1}{|\det A|}\, dx,
\end{aligned}$$

car $\mathcal{J}_{g^{-1}}(x)| = |\det A^{-1}| = |\det A|^{-1}$. Donc, la fonction

$$f_X(x) := f_Z(g^{-1}(x)) \frac{1}{|\det A|}, \quad x \in \mathbb{R}^n,$$

est une densité pour X. On trouve

$$\begin{aligned}
f_X(x) &= \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det A|} \exp\left(- \frac{\|A^{-1}(x-\mu)\|^2}{2} \right) \\
&= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det C}} \exp\left(- \frac{\langle x-\mu, C^{-1}(x-\mu)\rangle}{2} \right)
\end{aligned}$$

car

$$\begin{aligned}
\|A^{-1}(x-\mu)\|^2 &= \langle A^{-1}(x-\mu), A^{-1}(x-\mu)\rangle \\
&= \langle (x-\mu), (A^{-1})^t A^{-1}(x-\mu)\rangle \\
&= \langle (x-\mu), (A\cdot A^t)^{-1}(x-\mu)\rangle \\
&= \langle (x-\mu), C^{-1}(x-\mu)\rangle.
\end{aligned}$$

- Supposons C non inversible, alors A n'est pas inversible et

$$\det A = \sqrt{\det C} = 0.$$

Par conséquent,

$$H := g(\mathbb{R}^n) = \operatorname{Im} A + \mu$$

est un sous-espace affin de $\mathbb{R}^n$ de dimension $\dim H = \operatorname{rang} A < n$. En particulier,

$$\text{measure de Lebesgue de } H = \lambda_n(H) = 0.$$

Avec l'hypothèse que la loi de X soit absolument continue de densité f_X, on obtient une contradiction:

$$\begin{aligned}
1 = \mathbb{P}\{AZ+\mu \in H\} &= \mathbb{P}\{X \in H\} \\
&= \int\limits_{\mathbb{R}^n} \mathbb{1}_H(x) f_X(x)\, dx \\
&= \int\limits_{H} f_X(x)\, \lambda_n(dx) = 0.
\end{aligned}$$

Donc, la loi de X n'est pas absolument continue. $\square$

NOTATION 4.14. Tout vecteur gaussien dont la matrice de variance-covariance n'est pas inversible est dit *dégénéré*. La loi d'un tel vecteur est appelée *loi gaussienne dégénérée*.

DÉFINITION 4.15 (fonction caractéristique). Soit X un vecteur aléatoire de $\mathbb{R}^m$. On appelle *fonction caractéristique* de X la fonction

$$\varphi_X \colon \mathbb{R}^m \to \mathbb{C},$$
$$\varphi_X(a) := \mathbb{E}[e^{i\langle a,X\rangle}] = \mathbb{E}[\cos\langle a, X\rangle] + i\mathbb{E}[\sin\langle a, X\rangle].$$

On note que

$$\varphi_X(a) = \varphi_{\langle a,X\rangle}(1), \quad a \in \mathbb{R}^m.$$

EXEMPLE 4.16. Soit $X \sim N(\mu, C)$, c.à.d. pour tout $a \in \mathbb{R}^m$,

$$\langle a, X\rangle \sim N(\langle a,\mu\rangle, \langle a, Ca\rangle).$$

Alors

$$\varphi_X(a) = \varphi_{\langle a,X\rangle}(1) = e^{i\langle a,\mu\rangle - \frac{1}{2}\langle a,Ca\rangle}, \quad a \in \mathbb{R}^m.$$

REMARQUE 4.17 (Convergence en loi; *voir cours de Probabilité et Statistiques* 2). Soit $(X^n)_{n\geqslant 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^m$ et soit X une variable aléatoire continue à valeurs dans $\mathbb{R}^m$. On note

$$F_{X^n}(a) = \mathbb{P}\{X_1^n \leqslant a_1, \ldots, X_m^n \leqslant a_m\}$$

la fonction de répartition de X^n, et F_X la fonction de répartition de X.

Les conditions suivantes sont équivalentes:

(1) $\varphi_{X^n}(a) \to \varphi_X(a)$ pour tout $a \in \mathbb{R}^m$.
(2) $\mathbb{E}[h(X^n)] \to \mathbb{E}[h(X)]$ pour toute fonction $h\colon \mathbb{R}^m \to \mathbb{R}$ continue et borneé.
(3) $F_{X^n}(a) \to F_X(a)$ pour tout $a \in \mathbb{R}^m$.

Si une des ces conditions est vérifiée, on dit que $(X^n)_{n\geqslant 1}$ *converge en loi vers* X, et l'on note

$$X^n \xrightarrow{\mathscr{L}} X.$$

COROLLAIRE 4.18. *Les conditions suivantes sont equivalentes:*

(1) $X^n \xrightarrow{\mathscr{L}} X$ *dans* $\mathbb{R}^m$*;*
(2) *Pour tout* $a \in \mathbb{R}^m$ *on a la convergence* $\langle a, X^n\rangle \xrightarrow{\mathscr{L}} \langle a, X\rangle$ *dans* $\mathbb{R}$.

DÉFINITION 4.19. Soient $X_1, \ldots, X_n$ indépendantes et de loi $N(0,1)$. La loi de

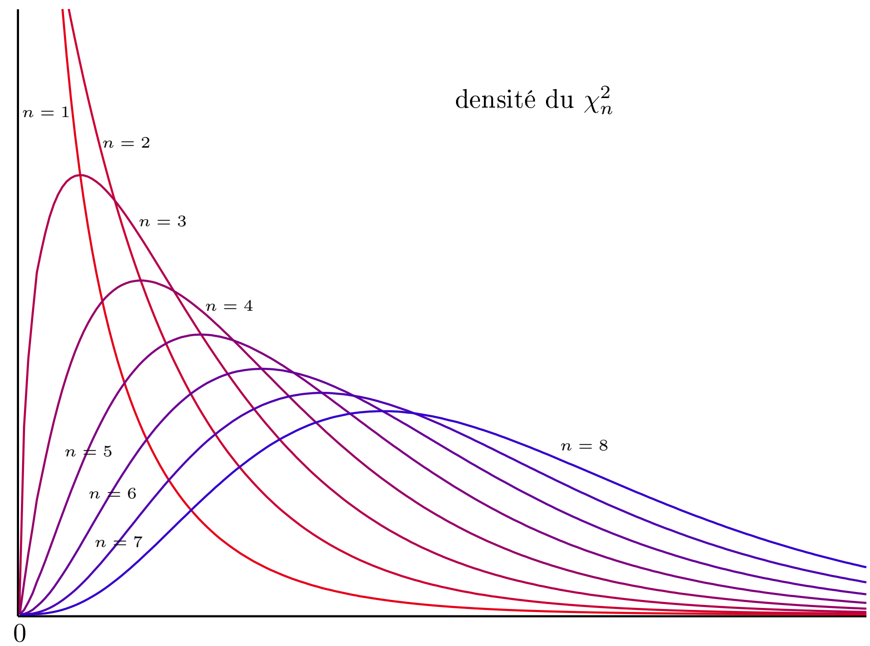
$$U := X_1^2 + \ldots + X_n^2$$

s'appelle *loi du chi-carré à n-degrés de liberté*. On la note χ_n^2.

EXEMPLE 4.20. On a

$$f_{\chi_n^2}(t) = \frac{t^{n/2-1}}{2^{n/2}\Gamma(n/2)}\, e^{-t/2}\, \mathbb{1}_{\mathbb{R}_+}(t), \quad t \in \mathbb{R}.$$

Preuve par récurrence.



4.3. Théorème de Cochran

THÉORÈME 4.1 (de Cochran). *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^m$ de loi $N(0, \sigma^2 \mathbb{1}_m)$ avec $\sigma \neq 0$. Soient $E_1, \dots, E_r$ des sous-espaces vectoriels de $\mathbb{R}^m$, deux à deux orthogonaux,* $\dim E_i \geqslant 1$ *pour tout i, tels que*

$$\mathbb{R}^m = \bigoplus_{i=1}^{r} E_i.$$

Notons P_i la matrice de la projection orthogonale sur E_i (dans la base canonique), $i = 1, \dots, r$. Alors

(1) $P_i X$, $i = 1, \dots, r$, *sont des vecteurs gaussiens mutuellement indépendants;*

(2) $P_i X \sim N(0, \sigma^2 P_i)$ *pour tout $i = 1, \dots, r$;*

(3) $\dfrac{\|P_i X\|^2}{\sigma^2} \sim \chi^2_{\dim E_i}$ *pour tout $i = 1, \dots, r$.*

PREUVE. Soit $d_i := \dim E_i$ pour $i = 1, \dots, r$ et $d_0 := 0$. On note $m_i := \sum_{k=0}^{i-1} d_k$ pour $i = 1, \dots, r$, et on fixe une base orthonormée $(v_1, \dots, v_m)$ de $\mathbb{R}^m$ adaptée à la décomposition

$$\mathbb{R}^m = \bigoplus_{i=1}^{r} E_i.$$

Autrement dit, pour tout $1 \leqslant i \leqslant r$, la famille

$$(v_{m_i+1}, \dots, v_{m_{i+1}})$$

est une base orthonormée de E_i.

Posons $Y_j := \langle v_j, X\rangle$. Alors

$$Y := (Y_1, \dots, Y_m) = P^t X$$

où P est la matrice du passage de la base canonique de $\mathbb{R}^m$ à la base $(v_1, \dots, v_m)$. Donc,

$$Y \sim N(0, P^t(\sigma^2 \mathbb{1}_m)P) = N(0, \sigma^2 \mathbb{1}_m),$$

c.à.d. les variables aléatoires $Y_1, \dots, Y_m$ sont i.i.d. de loi $N(0, \sigma^2)$. Par conséquent, les variables

$$P_i X = \sum_{j=m_i+1}^{m_{i+1}} Y_j v_j, \quad 1 \leqslant i \leqslant r,$$

sont indépendantes, et

$$P_i X \sim N(0, \underbrace{P_i(\sigma^2 \mathbb{1}_m)P_i^t}_{\sigma^2 P_i}),$$

car $P_i \mathbb{1}_m P_i^t = P_i P_i^t = P_i^2 = P_i$. Finalement, comme $v_1, \dots, v_m$ est une base orthonormée de $\mathbb{R}^m$, on obtient

$$\frac{\|P_i X\|^2}{\sigma^2} = \sum_{j=m_i+1}^{m_{i+1}} \frac{Y_j^2}{\sigma^2}.$$

Les variables

$$\frac{Y_1}{\sigma}, \dots, \frac{Y_m}{\sigma}$$

sont i.i.d. et suivent la loi $N(0,1)$. Donc la loi de $\|P_i X\|^2/\sigma^2$ est chi-carré à

$$m_{i+1} - m_i = d_i = \dim E_i$$

degrés de liberté. $\square$

Théorème 4.2 (Théorème de la limite central; cas multi-dimensionnel). *Soient $X_1, X_2, \dots$ des variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^m$ telles que $\mathbb{E}[X_1] = 0$ et $\mathbb{E}[\|X_1\|^2] < \infty$. Notons*

$$S_n := \sum_{i=1}^n X_i.$$

Alors

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathscr{L}} N(0, C), \quad \text{quand } n \to \infty,$$

où $C = \operatorname{var}(X_1)$, c.à.d. $C_{\ell k} = \mathbb{E}[X_{1,\ell}\, X_{1,k}]$ pour $1 \leqslant \ell, k \leqslant m$.

Preuve. Soit $a \in \mathbb{R}^m$ et $\hat{X}_i := \langle a, X_i\rangle$. Alors les $(\hat{X}_i)$ sont i.i.d. de moyenne $\mathbb{E}[\hat{X}_i] = 0$ et de variance égale à $\operatorname{var}(\hat{X}_i) = \langle a, Ca\rangle = \sigma^2$. Soit

$$\hat{S}_n := \sum_{i=1}^n \hat{X}_i.$$

Alors $\operatorname{var}(\hat{S}_n) = n\langle a, Ca\rangle = n\sigma^2$. Donc (par le cas uni-dimensionnel du Théorème limite central)

$$\frac{\hat{S}_n}{\sqrt{\operatorname{var}(\hat{S}_n)}} = \frac{\hat{S}_n}{\sqrt{n}\sigma} \xrightarrow{\mathscr{L}} N(0,1),$$

c.à.d. $\hat{S}_n/\sqrt{n} \xrightarrow{\mathscr{L}} N(0,\sigma^2)$. Par conséquent,

$$\frac{\langle a, S_n\rangle}{\sqrt{n}} \xrightarrow{\mathscr{L}} N(0, \langle a, Ca\rangle) \quad \forall a \in \mathbb{R}^m,$$

autrement dit $S_n/\sqrt{n} \xrightarrow{\mathscr{L}} N(0,C)$, car $\phi_{S_n/\sqrt{n}}(a) = \phi_{\langle a,S_n\rangle/\sqrt{n}}(1)$. □

4.4. Le test du χ^2

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace de probabilité et

$$\Omega = C_1 \,\dot{\cup} \ldots \dot{\cup}\, C_m$$

une décomposition de Ω où $C_i \in \mathscr{A}$ et $m \geqslant 2$. Soient $Y_1, \ldots, Y_n$ i.i.d. sur $(\Omega, \mathscr{A}, \mathbb{P})$ à valeurs dans $\{1, \ldots, m\}$ telles que

$$\mathbb{P}\{Y_i = k\} = \mu_k \text{ (inconnue)}, \quad k = 1, \ldots, m.$$

Considérons

$$N_k^n := \sum_{i=1}^n \mathbb{1}_{\{Y_i = k\}}, \quad k = 1, \ldots, m.$$

Soit p une distribution de probabilité sur $\{1, \ldots, m\}$:

$$p_k > 0, \quad k = 1, \ldots, m \quad (p_k \mathrel{\widehat{=}} \text{“probabilité de } C_k\text{”}).$$

On veut tester l'hypothèse

$$(\mathrm{H}_0) \quad \boxed{\mu = p} \qquad \text{contre} \qquad (\mathrm{H}_1) \quad \boxed{\mu \neq p}$$

Heuristique. Sous l'hypothèse H_0 on a

$$\mathbb{E}[N_k^n] = np_k.$$

Donc, on va rejeter H_0 si les déviations $N_k^n - np_k$ sont “assez grandes”.

Considérons la “statistique”

$$T_n := \sum_{k=1}^m \frac{(N_k^n - np_k)^2}{np_k} \quad \textit{(Pearson's sample function).}$$

Idée: Si la valeur de T_n est “assez grande”, on rejette H_0.

THÉORÈME 4.3 (Pearson, 1900). *Sous l'hypothèse H_0:*

$$\text{“}Y_1, \ldots, Y_n \textit{ i.i.d. de loi } p\text{”},$$

on a

$$T_n \xrightarrow{\mathscr{L}} \chi^2_{m-1}, \quad \textit{quand } n \to \infty.$$

PREUVE. (1) Sous l'hypothèse H_0,

$$(N_1^n, \ldots, N_m^n)$$

suit la *loi multinomiale* de paramètres $(n; p_1, \ldots, p_m)$, c.à.d.

$$\mathbb{P}\{N_1^n = n_1, \ldots, N_m^n = n_m\} = \frac{n!}{n_1! \cdot \ldots \cdot n_m!}\, p_1^{n_1} \cdot \ldots \cdot p_m^{n_m},$$

pour tout $n_1, \ldots, n_m \in \mathbb{N}$ tels que $n_1 + \ldots + n_m = n$. On remarque que

$$\frac{n!}{n_1! \cdot \ldots \cdot n_m!} = \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \ldots \cdot \binom{n-n_1-n_2-\ldots-n_{m-1}}{n_m}.$$

(2) On note

$$\sqrt{p} := (\sqrt{p_1}, \ldots, \sqrt{p_m}) \quad (\Rightarrow \|\sqrt{p}\| = 1)$$

et pour $x, y \in \mathbb{R}^m$:

$$x \otimes y \in \mathrm{Matr}(m \times m; \mathbb{R}), \quad (x \otimes y)_{k\ell} := x_k y_\ell.$$

Alors

$$Z_n := \left(\frac{N_1^n - np_1}{\sqrt{np_1}}, \ldots, \frac{N_m^n - np_m}{\sqrt{np_m}} \right) \xrightarrow{\mathscr{L}} N(0, C)$$

où

$$C := \mathbb{1}_m - \sqrt{p} \otimes \sqrt{p}.$$

En effet, sous H_0,

$$\begin{aligned} \frac{N_k^n - n\mathbb{P}\{Y_1 = k\}}{\sqrt{np_k}} &= \frac{1}{\sqrt{np_k}} \sum_{i=1}^n \left(\mathbb{1}_{\{Y_i = k\}} - \mathbb{E}[\mathbb{1}_{\{Y_i=k\}}] \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\mathbb{1}_{\{Y_i=k\}} - \mathbb{E}[\mathbb{1}_{\{Y_i=k\}}]}{\sqrt{p_k}} \right). \end{aligned}$$

Donc,

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(X^{(i)} - \mathbb{E}[X^{(i)}] \right) \text{ avec } X^{(i)} = \left(\frac{\mathbb{1}_{\{Y_i=1\}}}{\sqrt{p_1}}, \ldots, \frac{\mathbb{1}_{\{Y_i=m\}}}{\sqrt{p_m}} \right).$$

Par le *Théorème limite central* (Théorème 4.2) on obtient:

$$Z_n \xrightarrow{\mathscr{L}} Z, \quad \text{quand } n \to \infty,$$

avec $Z \sim N(0, C)$ où

$$\begin{aligned} C_{k\ell} &= \operatorname{cov}\left(\frac{\mathbb{1}_{\{Y_1=k\}}}{\sqrt{p_k}}, \frac{\mathbb{1}_{\{Y_1=\ell\}}}{\sqrt{p_\ell}} \right) \\ &= \frac{1}{\sqrt{p_k p_\ell}} \Big(\mathbb{P}\{Y_1 = k, Y_1 = \ell\} - \underbrace{\mathbb{P}\{Y_1 = k\}}_{=p_k} \underbrace{\mathbb{P}\{Y_1 = \ell\}}_{=p_\ell} \Big) \\ &= \begin{cases} -\sqrt{p_k p_\ell}, & \text{si } k \neq \ell \\ 1 - p_k, & \text{si } k = \ell \end{cases} \\ &= \left(\mathbb{1}_m - \sqrt{p} \otimes \sqrt{p} \right)_{k\ell}. \end{aligned}$$

(3) Rappelons que Z_n et Z sont à valeurs dans $\mathbb{R}^m$ et que $Z_n \xrightarrow{\mathscr{L}} Z$. Donc avec

$$h\colon \mathbb{R}^m \to \mathbb{R}_+, \quad x \longmapsto \|x\|^2,$$

on obtient

$$h(Z_n) \xrightarrow{\mathscr{L}} h(Z), \quad n \to \infty,$$

c.à.d.

$$T_n = \|Z_n\|^2 \xrightarrow{\mathscr{L}} \|Z\|^2, \quad n \to \infty.$$

Il reste à déterminer la loi de $D := \|Z\|^2$. On pose

$$v := \sqrt{p} \quad \text{et} \quad A := \sqrt{p} \otimes \sqrt{p}.$$

Alors on trouve

$$Ax = \langle x, v \rangle \, v$$

pour tout $x \in \mathbb{R}^m$. La matrice A est donc la matrice de la projection orthogonale sur $E_1 := \mathrm{span}(v)$, et par conséquent,

$$C = \mathbb{1}_m - A$$

est la matrice de la projection orthogonale sur $E_2 := E_1^\perp$ $(\dim E_2 = m - 1)$.

Par le Théorème de Cochran on obtient:

(a) Z a même loi que CU avec $U \sim N(0, \mathbb{1}_m)$, et
(b) $D = \|Z\|^2$ a même loi que $\|CU\|^2 \sim \chi^2_{\dim E_2} = \chi^2_{m-1}$. □

Exemple 4.21 (Expérience de Gregor Johann Mendel; 8 février 1856). On croise 2 populations (pures) de pois, et on s'intéresse aux caractères "*couleur*" et "*forme*". On a

$$\text{“\textit{couleur}”} \qquad \begin{cases} \text{C (jaune),} & \text{dominant} \\ \text{c (vert),} & \text{récessif} \end{cases}$$

$$\text{“\textit{forme}”} \qquad \begin{cases} \text{R (rond),} & \text{dominant} \\ \text{r (ridé),} & \text{récessif.} \end{cases}$$

Mendel découvre qu'un caractère peut présenter deux formes différentes (aujourd'hui appelées *allèles* ou *gènes homologues*). Un organisme hérite de deux facteurs pour chaque caractère (les facteurs héréditaires de Mendel sont aujourd'hui appelés «*allèles*»). Le facteur dominant masque le facteur récessif. Mendel a noté le facteur dominant à l'aide d'une majuscule et l'autre, le récessif, à l'aide de la même lettre mais en minuscule.

En croisant deux individus de génotype CcRr, il y a 16 génotypes équiprobables pour les descendants:

CCRR	$\longrightarrow$	jaune, ronde
CcRR	$\longrightarrow$	jaune, ronde
cCRR	$\longrightarrow$	jaune, ronde
ccRR	$\longrightarrow$	vert, ronde
CCRr	$\longrightarrow$	jaune, ronde
CcRr	$\longrightarrow$	jaune, ronde
cCRr	$\longrightarrow$	jaune, ronde
ccRr	$\longrightarrow$	vert, ronde
CCrR	$\longrightarrow$	jaune, ronde
CcrR	$\longrightarrow$	jaune, ronde
cCrR	$\longrightarrow$	jaune, ronde
ccrR	$\longrightarrow$	vert, ronde
CCrr	$\longrightarrow$	jaune, ridé
Ccrr	$\longrightarrow$	jaune, ridé
cCrr	$\longrightarrow$	jaune, ridé
ccrr	$\longrightarrow$	vert, ridé

Les phénotypes devraient être distribués de la façon suivante:

	jaune, rond	jaune, ridé	vert, rond	vert, ridé
Fréquence théorique	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
Effectifs ($n = 556$)	315	101	108	32
Fréquence empirique	$\frac{315}{556}$	$\frac{101}{556}$	$\frac{108}{556}$	$\frac{32}{556}$

H_0: *les fréquences d'apparition des différents caractères sont bien données par les prédictions de Mendel*:

$$\frac{9}{16}, \quad \frac{3}{16}, \quad \frac{3}{16}, \quad \frac{1}{16}.$$

Alors

$$T_{556} = \frac{\left(315 - 556 \times \frac{9}{16}\right)^2}{556 \times \frac{9}{16}} + \ldots + \frac{\left(32 - 556 \times \frac{1}{16}\right)^2}{556 \times \frac{1}{16}} \cong 0{,}47.$$

Pour un seuil de 5%, on obtient

$$\mathbb{P}_{\mathrm{H}_0}\{\chi_3^2 > C\} = 0{,}05 \quad \text{avec } C \approx 7{,}82.$$

Puisque $0,47 \ll 7{,}82$ on ne rejette pas l'hypothèse H_0.