

AGRÉGATION OPTION A

CHRISTOPHE POQUET, MARIELLE SIMON

CONTENTS

1. Martingales	1
1.1. Définitions	1
1.2. Temps d'arrêt	2
1.3. Théorème d'arrêt	4
1.4. Inégalité maximale de Doob	4
1.5. Convergences	4
2. Chaînes de Markov	5
2.1. Définitions	5
2.2. Exemples	6
2.3. Propriété de Markov	8
2.4. Classification des états	9
2.5. Mesures invariantes	11
2.6. Récurrence positive	12
2.7. Théorèmes limites	13
2.8. Quelques exemples	13
References	15

1. MARTINGALES

Les martingales sont utilisées pour modéliser les jeux de hasard équitables : si, à un instant t , on a perdu ou gagné une certaine somme S , alors on n'a pas plus de chance dans le futur d'augmenter ou de diminuer le gain. Un théorème dit que si un joueur début le jeu avec une fortune initiale finie, il n'existe pas de stratégie pour gagner à coup sûr (et donc les casinos s'enrichissent).

Les martingales ont d'importantes propriétés qui permettent d'obtenir des résultats de convergence forts.

1.1. Définitions.

Définition 1.1 (Filtration). *Une filtration d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est une suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous-tribus de \mathcal{F} . On dit que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ est un espace de probabilité filtré.*

Remarque 1.2. *Souvent on utilise la convention $\mathcal{F}_{-1} := \{\emptyset, \Omega\}$.*

Définition 1.3 (Processus adapté). *Un processus $(X_n)_{n \in \mathbb{N}}$ est dit adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si, pour tout $n \in \mathbb{N}$, X_n est \mathcal{F}_n -mesurable.*

Exemple 1.4. *Un processus (X_n) est toujours adapté pour sa filtration canonique définie par*

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n).$$

Définition 1.5 (Martingale). *Le processus $(X_n)_{n \in \mathbb{N}}$ est une martingale pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si les trois conditions suivantes sont vérifiées :*

- (i) pour tout $n \in \mathbb{N}$, $\mathbb{E}[|X_n|] < \infty$ (i.e. X_n est dans L^1)
- (ii) $(X_n)_{n \in \mathbb{N}}$ est adapté à $(\mathcal{F}_n)_{n \in \mathbb{N}}$
- (iii) pour tout $n \in \mathbb{N}$,

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n \quad \text{p.s.}$$

On dit que (X_n) est une sur-martingale (resp. sous-martingale) si (i) et (ii) sont vérifiées et si (iii) est remplacé par

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n \quad \text{p.s.} \quad \text{resp.} \quad \mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n \quad \text{p.s.}$$

Exemple 1.6 (Quelques exemples). *Voici quelques exemples à connaître :*

- (1) Soit (\mathcal{F}_n) une filtration, et $X \in L^1(\Omega)$ une variable aléatoire intégrable. Alors $M_n = \mathbb{E}[X | \mathcal{F}_n]$ est une martingale pour (\mathcal{F}_n) .
- (2) Soient $(Z_\ell)_{\ell \in \mathbb{N}}$ des v.a.i.i.d. intégrables, d'espérance nulle. Alors

$$M_n = Z_0 + \dots + Z_n$$

est une martingale pour sa filtration canonique. En effet:

$$\begin{aligned} \mathbb{E}[M_{n+1} | \mathcal{F}_n] &= \underbrace{\mathbb{E}[M_n | \mathcal{F}_n]}_{=M_n \text{ car mesurable}} + \underbrace{\mathbb{E}[X_{n+1} | \mathcal{F}_n]}_{=\mathbb{E}[X_{n+1}] = 0 \text{ car } \perp} \\ &= M_n \end{aligned}$$

Proposition 1.7. (1) Si (X_n) est une martingale pour la filtration (\mathcal{F}_n) alors $\mathbb{E}[X_n] = \mathbb{E}[X_0]$.

De plus, $\mathbb{E}[X_{n+k} | \mathcal{F}_n] = X_n$ pour tout $k \geq 0$.

- (2) Si (X_n) est une sur-martingale (resp. sous-martingale) alors $\mathbb{E}[X_n] \leq \mathbb{E}[X_0]$ (resp. $\mathbb{E}[X_n] \geq \mathbb{E}[X_0]$).

Proposition 1.8. Si (X_n) est une martingale, si $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est convexe, et si, pour tout n , $\varphi(X_n)$ est intégrable, alors $(\varphi(X_n))_{n \in \mathbb{N}}$ est une sous-martingale.

Exemple 1.9. Par exemple, si (X_n) est une martingale, alors $(|X_n|)$, (X_n^2) , et $(X_n^+ = \max(X_n, 0))$ sont des sous-martingales.

1.2. Temps d'arrêt. On souhaite arrêter un processus à des temps aléatoires, dépendant uniquement du passé et du présent (pas du futur). On rappelle ici la définition :

Définition 1.10 (Temps d'arrêt). *Une application mesurable $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ est un temps d'arrêt pour la filtration (\mathcal{F}_n) si, pour tout $n \in \mathbb{N}$,*

$$\{T = n\} \in \mathcal{F}_n \tag{1}$$

Remarque 1.11. L'événement $\{T = \infty\}$ appartient à $\mathcal{F}_\infty := \sigma(\cup_{n \in \mathbb{N}} \mathcal{F}_n)$.

Remarque 1.12. Il est important de remarquer que la condition (1) est équivalente à $\{T \leq n\} \in \mathcal{F}_n$ pour tout n . Remarquons aussi : $\{T < n\} = \{T \leq n-1\} \in \mathcal{F}_{n-1}$.

Exemple 1.13. Voici quelques exemples :

- (1) Un temps déterministe : $T(\omega) \equiv n_0$ est un temps d'arrêt.
- (2) Les temps d'atteinte sont des temps d'arrêt pour la filtration naturelle des processus, i.e.

$$T_B = \inf \{n \in \mathbb{N} ; X_n \in B\}$$

car

$$\{T_B = n\} = \bigcup_{k=0}^{n-1} \{X_k \notin B\} \cup \{X_n \in B\}$$

(3) Un temps défini comme $\tau_B = \sup\{n \in \mathbb{N} ; X_n \in B\}$ n'est pas un temps d'arrêt, car il dépend de tout le futur.

Proposition 1.14. Si S et T sont deux temps d'arrêt pour la même filtration (\mathcal{F}_n) , alors $S+T$, $S \wedge T$ et $S \vee T$ sont des temps d'arrêt pour (\mathcal{F}_n) .

Définition 1.15 (Tribu des événements antérieurs). Si T est un temps d'arrêt, on définit

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty ; \text{pour tout } n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}$$

la tribu des événements antérieurs à T .

Exemple 1.16. Soit $(X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans \mathbb{Z} et soit (\mathcal{F}_n) sa filtration canonique. Notons T le temps d'atteinte de 0. On considère A l'événement :

$$A = \{ \text{la trajectoire } (X_n) \text{ passe par un nombre pair} \}$$

Alors on peut récrire $A = \{\tau < \infty\}$ où τ est le temps d'atteinte de l'ensemble des entiers pairs. Alors

$$A \in \mathcal{F}_T$$

En effet : comme $\tau \leq T$ p.s., on a

$$\{\tau < \infty\} \cap \{T = n\} = \{\tau \leq n\} \cap \{T = n\} \in \mathcal{F}_n$$

Proposition 1.17. Si S et T sont deux temps d'arrêt pour la même filtration (\mathcal{F}_n) et si $S \leq T$ alors $\mathcal{F}_S \subset \mathcal{F}_T$.

Proposition 1.18. La variable aléatoire $\mathbf{1}_{T < \infty} X_T$ est \mathcal{F}_T -mesurable.

Proof.

$$\{\mathbf{1}_{T < \infty} X_T \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\} \in \mathcal{F}_n$$

□

On parle maintenant du processus arrêté.

Proposition 1.19. Si $(X_n)_{n \in \mathbb{N}}$ est une martingale (resp. sur-martingale, resp. sous-martingale) pour (\mathcal{F}_n) et si T est un temps d'arrêt pour (\mathcal{F}_n) alors $(X_{n \wedge T})_{n \in \mathbb{N}}$ est une martingale (resp. sur-martingale, resp. sous-martingale) pour (\mathcal{F}_n) .

Proof. On a

$$\begin{aligned} \mathbb{E}[X_{(n+1) \wedge T} - X_{n \wedge T} | \mathcal{F}_n] &= \mathbb{E}[(X_{(n+1) \wedge T} - X_{n \wedge T}) \mathbf{1}_{T > n} + 0 \times \mathbf{1}_{T \leq n} | \mathcal{F}_n] \\ &= \mathbb{E}[(X_{n+1} - X_n) \underbrace{\mathbf{1}_{T > n}}_{\text{mesurable}} | \mathcal{F}_n] \\ &= \mathbf{1}_{T > n} \mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] = 0. \end{aligned}$$

□

1.3. Théorème d'arrêt. D'après le résultat précédent, si T est un temps d'arrêt, et si (X_n) est une martingale (pour (\mathcal{F}_n)) alors

$$\mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_0]$$

De plus, si $T < \infty$ p.s., alors $X_{T \wedge n} \xrightarrow[n \rightarrow \infty]{p.s.} X_T$. Si on pouvait passer à la limite sous l'espérance, alors on aurait $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, ce qui a d'importantes conséquences pratiques. On va énoncer le résultat général :

Théorème 1.20 (Théorème d'arrêt). *Si l'une des trois conditions suivantes est satisfaite :*

- (1) T est borné (autrement dit, il existe $N \in \mathbb{N}$ tel que $T \leq N$ p.s.)
- (2) T est intégrable (i.e. $\mathbb{E}[T] < \infty$) et il existe $M > 0$ tel que, pour tout $n \geq 0$, $|X_{n+1} - X_n| \leq M$ p.s.
- (3) $T < \infty$ p.s. et $(X_{T \wedge n})$ est uniformément borné (il existe $C > 0$ tel que, pour tout n , $X_{T \wedge n} \leq C$ p.s.),

alors $(X_{T \wedge n})$ est intégrable, et on a l'égalité : $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Remarque 1.21. *Si l'une des trois conditions est satisfaite et si (X_n) est une sur-(resp. sous-) martingale, alors $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ (resp. $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$).*

Exemple 1.22. *Donnons un contre-exemple au théorème d'arrêt. Supposons $X_0 = 1$ p.s., et $X_{n+1} = X_n \times Z_{n+1}$ où les (Z_i) sont i.i.d. de même loi $\frac{1}{2}\delta_2 + \frac{1}{2}\delta_0$. On note T le temps d'atteinte de 0. Alors T est fini presque sûrement (il suit la loi géométrique de paramètre $\frac{1}{2}$). Mais on a $\mathbb{E}[X_0] = 1$ et $\mathbb{E}[X_T] = 0$.*

1.4. Inégalité maximale de Doob.

Théorème 1.23 (Inégalité maximale de Doob). *Si (X_n) est une sous-martingale positive, alors, pour tout $\lambda > 0$,*

$$\mathbb{P}\left(\max_{0 \leq k \leq n} X_k > \lambda\right) \leq \frac{\mathbb{E}[X_n]}{\lambda}$$

En particulier, si (X_n) est une martingale et si $X_n \in L^p$ pour un certain $p \geq 1$, alors on peut appliquer ce résultat à $(|X_n|^p)$ qui est une sous-martingale positive, et on obtient :

$$\mathbb{P}\left(\max_{0 \leq k \leq n} |X_k| > \lambda\right) = \mathbb{P}\left(\max_{0 \leq k \leq n} |X_k|^p > \lambda^p\right) \leq \frac{\mathbb{E}[|X_n|^p]}{\lambda^p}$$

Exemple 1.24. *On applique ce résultat à la marche aléatoire simple sur \mathbb{Z} : $X_n = Z_1 + \dots + Z_n$ où les Z_i sont i.i.d. de même loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Alors, $\mathbb{E}[X_n^2] = \text{Var}(X_n) = n\text{Var}(Z_1) = n$ et*

$$\mathbb{P}\left(\max_{0 \leq k \leq n} |X_k| \geq n^{\frac{1}{2} + \varepsilon}\right) \leq \frac{\mathbb{E}[X_n^2]}{n^{1+2\varepsilon}} = n^{-2\varepsilon}$$

1.5. Convergences.

Théorème 1.25 (Martingales bornées dans L^1). *Si (X_n) est une martingale (resp. sous- ou sur-martingale) vérifiant $\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < \infty$, alors*

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} X_\infty, \quad \text{où } X_\infty \text{ est une v.a. intégrable.}$$

Remarque 1.26. *La convergence a lieu p.s. mais pas dans L^1 . E, effet, prenons le même exemple que tout à l'heure : $X_0 = 1$ p.s., et $X_{n+1} = X_n \times Z_{n+1}$ où les (Z_i) sont i.i.d. de même loi $\frac{1}{2}\delta_2 + \frac{1}{2}\delta_0$. Alors $\mathbb{E}[X_n] = \mathbb{E}[X_0] = 1$, mais $X_n \rightarrow 0$ p.s.*

Théorème 1.27 (Martingales bornées dans L^2). *Si (X_n) est une martingale uniformément bornée dans L^2 i.e. vérifiant $\sup_{n \in \mathbb{N}} \mathbb{E}[X_n^2] < \infty$ alors*

$$X_n \xrightarrow[n \rightarrow \infty]{p.s. \text{ et dans } L^2} X_\infty, \quad \text{où } X_\infty \text{ est une v.a. intégrable.}$$

De plus : pour tout $n \in \mathbb{N}$

$$X_n = \mathbb{E}[X_\infty | \mathcal{F}_n].$$

Les martingales dans L^2 ont des propriétés très intéressantes, par exemple :

Théorème 1.28 (Pythagore pour les martingales). *Soit (X_n) une martingale telle que $\mathbb{E}[X_n^2] < \infty$ pour tout n . alors, pour tous $m \leq n \leq p$ entiers,*

$$\mathbb{E}[(X_n - X_m)(X_p - X_n)] = 0$$

et en particulier

$$\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \sum_{k=0}^{n-1} \mathbb{E}[(X_{k+1} - X_k)^2]$$

Proof. On a

$$\begin{aligned} \mathbb{E}[(X_n - X_m)(X_p - X_n)] &= \mathbb{E}[\mathbb{E}[(X_n - X_m)(X_p - X_n) | \mathcal{F}_n]] \\ &= \mathbb{E}[(X_n - X_m)\mathbb{E}[X_p - X_n | \mathcal{F}_n]] \\ &= \mathbb{E}[(X_n - X_m) \times 0] = 0. \end{aligned}$$

□

Théorème 1.29 (Convergence dans L^1). *Si (X_n) est une martingale, alors les conditions suivantes sont équivalentes :*

- (i) (X_n) converge vers une v.a. X_∞ p.s. et dans L^1 .
- (ii) il existe une v.a. $Y \in L^1$ intégrable telle que $X_n = \mathbb{E}[Y | \mathcal{F}_n]$ pour tout $n \in \mathbb{N}$.
- (iii) la suite (X_n) est uniformément intégrable, au sens suivant :

$$\lim_{c \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n| \mathbf{1}_{|X_n| > c}] = 0.$$

Exemple 1.30. *Si (X_n) vérifie $\sup_{n \in \mathbb{N}} |X_n| \leq M$ p.s., alors (X_n) est uniformément intégrable. C'est aussi vrai si*

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^{1+\alpha}] < \infty$$

pour un $\alpha > 0$.

2. CHAÎNES DE MARKOV

2.1. Définitions.

Définition 2.1. *Soit E un ensemble fini ou dénombrable.*

Une suite $X = (X_n)$ de variables aléatoires à valeurs dans E , définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est une chaîne de Markov de loi initiale μ_0 si la loi de X_0 est μ_0 et si, pour tout $n \geq 0$, pour tous $x_1, x_2, \dots, x_n, x_{n+1} \in E$,

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

dès lors que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$.

La chaîne est dite homogène si la probabilité de transition $\mathbb{P}(X_{n+1} = y | X_n = x) =: P(x, y)$ ne dépend pas de n . On appelle matrice de transition (ou noyau de transition) l'application $P : E \times E \rightarrow [0, 1]$, qui vérifie

$$\sum_{y \in E} P(x, y) = 1$$

À partir de maintenant on ne considère que des chaînes de Markov homogènes.

Proposition 2.2. *Soit X_0 une v.a. à valeurs dans $(E, \mathcal{P}(E))$, et soit (U_n) une suite de v.a. i.i.d. à valeurs dans un espace mesurable (G, \mathcal{G}) et indépendantes de X_0 . Enfin, soit $f : E \times F \rightarrow E$*

une application mesurable. La suite récurrente aléatoire

$$X_{n+1} = f(X_n, U_{n+1})$$

est une chaîne de Markov homogène à valeurs dans E .

Proof. On vérifie la propriété :

$$\begin{aligned} & \mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \\ &= \mathbb{P}(f(X_n, U_{n+1}) = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \\ &= \frac{\mathbb{P}(f(x_n, U_{n+1}) = x_{n+1}, X_0 = x_0, \dots, X_n = x_n)}{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)} \\ &= \mathbb{P}(f(x_n, U_{n+1}) = x_{n+1}) \quad \text{car } U_{n+1} \text{ est indépendante des } (X_i) \\ &= \mathbb{P}(f(x_n, U_1) = x_{n+1}) \quad \text{car les } U_i \text{ sont de même loi.} \end{aligned}$$

□

La loi d'une chaîne de Markov est entièrement donnée par sa loi initiale μ_0 , et sa matrice de transition P . On définit par récurrence les itérés de P comme suit : $P^0 := P$ et

$$P^{n+1}(x, y) = \sum_{z \in E} P(x, z)P^n(z, y) \quad (\text{produit standard de matrices dans le cas } E \text{ fini})$$

Il est facile de voir que pour tout $n \geq 1$, P^n est encore un noyau de transition.

On identifie les mesures ν sur E à des vecteurs lignes $(\nu(x_1), \dots, \nu(x_n), \dots)$ et les fonctions sur E à des vecteurs colonnes ${}^T(f(x_1), \dots, f(x_n), \dots)$. Les produits entre "matrice" et "vecteur" s'étendent aisément au cas E dénombrable. En particulier, si ν est une mesure sur E et $f : E \rightarrow \mathbb{R}$ est une fonction définie sur E , alors

$$\int_E f d\nu = \nu f = \sum_{x \in E} \nu(x) f(x), \quad \text{i.e. un produit scalaire entre deux vecteurs.}$$

Proposition 2.3. Soit (X_n) une chaîne de Markov homogène de probabilité de transition P et de loi initiale μ_0 . Alors, pour tout $x_0, x_1, \dots, x_n \in E$,

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)$$

De plus, la loi de X_n est $\mu_n = \mu_0 P^n$. En particulier,

$$\mathbb{P}(X_n = y | X_0 = x) = P^n(x, y) \quad \text{et} \quad \mathbb{E}[f(X_n)] = \mu_0 P^n f.$$

Proposition 2.4 (Relation de Chapman-Kolmogorov). Pour tout $n, m \geq 1$, pour tout $x, y \in E$

$$\begin{aligned} \mathbb{P}(X_{n+m} = y | X_0 = x) &= \sum_{z \in E} \mathbb{P}(X_m = y | X_0 = z) \mathbb{P}(X_n = z | X_0 = x) \\ &= \sum_{z \in E} P^m(z, y) P^n(x, z) \end{aligned}$$

2.2. Exemples.

2.2.1. Cas E fini.

- Graphe des transitions

- **Marche aléatoire sur les sommets d'un graphe.** Soit V un ensemble de sommets d'un graphe supposé connexe. On note $x \sim y$ (ou de manière équivalente $y \sim x$) s'il existe une arête (non orientée) reliant x à y . Pour tout $x \in V$, on note le nombre de ses voisins $d(x) = \#\{y \in V ; y \sim x\} \geq 1$ que l'on suppose fini. L'espace d'états est V , et la matrice de transition est donnée par

$$P(x, y) = \begin{cases} \frac{1}{d(x)} & \text{si } y \sim x \\ 0 & \text{sinon.} \end{cases}$$

- **Urne d'Ehrenfest.** L'espace d'états est $\{0, 1, \dots, d\}$ et la matrice de transition est donnée par

$$P(x, y) = \begin{cases} x/d & \text{si } y = x - 1 \\ (d - x)/d & \text{si } y = x + 1 \\ 0 & \text{sinon.} \end{cases}$$

- **Ruine du joueur.** L'espace d'états est $\{0, 1, \dots, K\}$ et la matrice de transition est donnée pour $p \in (0, 1)$ par

$$\begin{aligned} P(0, 0) &= 1 \\ P(K, K) &= 1 \\ P(x, x + 1) &= p \quad \text{et } P(x, x - 1) = 1 - p \quad \text{pour tout } x \in \{1, \dots, K - 1\} \end{aligned}$$

2.2.2. Cas E dénombrable.

- **Marche aléatoire simple symétrique sur \mathbb{Z}^d .** Soit $(e_i)_{i=1, \dots, d}$ est une base de \mathbb{Z}^d . Un marcheur fait un pas de longueur 1 toutes les secondes, il choisit sa direction au hasard uniformément sur $\{e_i, -e_i ; i = 1, \dots, d\}$. Il part de l'origine 0 et on note X_n sa position au temps n . La suite (X_n) est une chaîne de Markov sur \mathbb{Z}^d de matrice de transition

$$P(x, y) = \begin{cases} 1/(2d) & \text{si } |x - y| = 1 \\ 0 & \text{sinon.} \end{cases}$$

- **Chaîne de vie et de mort.** L'espace d'états est \mathbb{N} et la matrice de transition est définie par

$$P(x, x - 1) = q_x, \quad P(x, x + 1) = p_x, \quad P(x, x) = r_x,$$

où les p_x, q_x, r_x vérifient

$$q_0 = 0, \quad q_x + p_x + r_x = 1, \quad \begin{cases} p_x > 0 & \text{pour tout } x \geq 0 \\ q_x > 0 & \text{pour tout } x \geq 1 \end{cases}$$

2.3. Propriété de Markov. Soit $X = (X_n)$ une chaîne de Markov sur E de matrice de transition P et de loi initiale μ . On peut voir X comme une variable aléatoire à valeur dans l'espace des trajectoires $E^{\mathbb{N}}$. Pour cela, on munit $E^{\mathbb{N}}$ de la tribu \mathcal{B} engendrée par les cylindres $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$. Le résultat suivant est fondamental (et peut être admis) :

Théorème 2.5. *Il existe une unique mesure de probabilité \mathbb{P}_μ sur $(E^{\mathbb{N}}, \mathcal{B})$ telle que, sous \mathbb{P}_μ , (X_n) est une chaîne de Markov sur E de loi initiale μ et de noyau de transition P .*

À partir de maintenant on supposera que la chaîne de Markov est toujours définie sur l'espace de probabilité $(E^{\mathbb{N}}, \mathcal{B}, \mathbb{P}_\mu)$, au lieu de $(\Omega, \mathcal{F}, \mathbb{P})$ comme initialement, et sa matrice de transition

est P . On l'appelle alors la *chaîne de Markov canonique*. Lorsque $\mu = \delta_x$ où $x \in E$ on note plus simplement $\mathbb{P}_{\delta_x} = \mathbb{P}_x$. On note \mathbb{E}_μ l'espérance par rapport à \mathbb{P}_μ .

On introduit à présent la chaîne X^{+k} *décalée d'un certain rang k* , comme suit :

$$X_n^{+k} := X_{k+n}, \quad \text{pour tout } n \geq 0.$$

On la note aussi parfois $(\theta_k X)_n$, avec θ_k l'opérateur de translation.

Théorème 2.6 (Propriété de Markov faible). *Pour tout $A \in \mathcal{B}$, pour tout $k \in \mathbb{N}$*

$$\mathbb{P}_\mu(X^{+k} \in A | X_k = x_k, \dots, X_0 = x_0) = \mathbb{P}_{x_k}(X \in A)$$

dès lors que $\mathbb{P}_\mu(X_k = x_k, \dots, X_0 = x_0) > 0$. En particulier,

$$\begin{aligned} \mathbb{P}_\mu(X_{k+1} = y_1, \dots, X_{k+r} = y_r | X_k = x_k, \dots, X_0 = x_0) &= \mathbb{P}_{x_k}(X_1 = y_1, \dots, X_r = y_r) \\ &= P(x_k, y_1)P(y_1, y_2) \cdots P(y_{r-1}, y_r). \end{aligned}$$

De manière encore plus générale, on énonce la propriété de Markov faible de la manière suivante :

Théorème 2.7 (Propriété de Markov faible, énoncé plus général). *Soit $F : (E^{\mathbb{N}}, \mathcal{B}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ une fonction positive mesurable. Alors, quelque soit $\psi : E^{k+1} \rightarrow \mathbb{R}_+$ on a*

$$\mathbb{E}_\mu[\psi(X_0, \dots, X_k) F(X^{+k})] = \mathbb{E}_\mu[\psi(X_0, \dots, X_k) \mathbb{E}_{X_k}[F(X)]]$$

ce qui est équivalent à

$$\mathbb{E}_\mu[F(X^{+k}) | \mathcal{F}_k] = \mathbb{E}_{X_k}[F] \quad \mathbb{P}_\mu\text{-a.s.}$$

Cette propriété s'étend aux *temps aléatoires* qui possèdent certaines propriétés.

Définition 2.8 (Temps d'arrêt). *Un temps d'arrêt T est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$ telle que, pour tout $n \in \mathbb{N}$*

$$\{T = n\} \in \mathcal{F}_n = \sigma(X_0, \dots, X_n)$$

Théorème 2.9 (Propriété de Markov forte). *Soit T un temps d'arrêt. Alors, pour tout $k \in \mathbb{N}$, pour tout $A \in \mathcal{B}$,*

$$P_\mu(X^{+T} \in A | T = k, X_k = x_k, \dots, X_0 = x_0) = \mathbb{P}_{x_k}(X \in A)$$

ou encore : pour toute fonction $F : (E^{\mathbb{N}}, \mathcal{B}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ positive mesurable, et toute fonction $\psi : E^{k+1} \rightarrow \mathbb{R}_+$,

$$\mathbb{E}_\mu[\mathbf{1}_{T=k} \psi(X_0, \dots, X_k) F(X^{+T})] = \mathbb{E}_\mu[\mathbf{1}_{T=k} \psi(X_0, \dots, X_k) \mathbb{E}_{X_T}[F(X)]]$$

En particulier, définissons la tribu engendrée par le temps d'arrêt :

$$\mathcal{F}_T := \{A \in \mathcal{B} ; A \cap \{T = k\} \in \mathcal{F}_k \text{ pour tout } k \in \mathbb{N}\}$$

Alors, si $x, y \in E$ et T est un temps d'arrêt tel que, $T < \infty$ et $X_T = y$ \mathbb{P}_x -p.s., on peut dire que la chaîne translatée X^{+T} est une chaîne de Markov de matrice de transition P , de loi initiale δ_y , et est indépendante de \mathcal{F}_T .

2.4. Classification des états.

Définition 2.10. *Pour $x, y \in E$, on note $x \rightarrow y$ s'il existe $n \in \mathbb{N}$ tel que $P^n(x, y) > 0$. On dit que deux états $x, y \in E$ communiquent si $x \rightarrow y$ et $y \rightarrow x$. On note alors $x \leftrightarrow y$. On dit que x est absorbant si $P(x, x) = 1$.*

La relation \leftrightarrow est une relation d'équivalence sur E dont les classes sont appelées les classes de communication de la chaîne de Markov.

Une classe de communication C est fermée si, pour tout $x \in C$, il n'existe aucun $y \notin C$ tel que $x \rightarrow y$. S'il n'y a qu'une seule classe de communication $C = E$, la chaîne de Markov est dite irréductible.

Exemple 2.11. La marche aléatoire simple sur \mathbb{Z}^d est irréductible. La ruine du joueur possède deux états absorbants, et une classe de communication ouverte.

On s'intéresse aux états après une longue trajectoire de la chaîne de Markov : en particulier, quels états ne seront plus visités après un certain temps, et quels sont ceux qui au contraire seront revisités perpétuellement ?

2.4.1. *Réurrence et transience.* Soit $x \in E$ et (X_n) une chaîne de Markov sur E de noyau P . On définit

$$T_x := \inf \{n > 0 ; X_n = x\}$$

à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$.

Définition 2.12. On dit que x est transient si $\mathbb{P}_x(T_x < \infty) < 1$ et on dit que x est récurrent si $\mathbb{P}_x(T_x < \infty) = 1$.

Un état récurrent est dit récurrent nul si $\mathbb{E}_x[T_x] = \infty$ et récurrent positif si $\mathbb{E}_x[T_x] < \infty$.

Ainsi, si x est récurrent, alors partant de x , la chaîne de Markov revient presque sûrement en x . Par la propriété de Markov forte au temps T_x , un second retour en x est aussi presque sûr, etc. Pour étudier cela, on introduit le nombre de visites en x :

$$N_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}}$$

Proposition 2.13. Soit $x \in E$.

- (1) L'état x est récurrent si et seulement si $\mathbb{P}_x(N_x = \infty) = 1$.
- (2) L'état x est transient si et seulement si $\mathbb{P}_x(N_x < \infty) = 1$. De plus N_x suit une loi géométrique de paramètre $p = \mathbb{P}_x(T_x = \infty)$ et

$$\mathbb{E}_x[N_x] = \frac{1}{\mathbb{P}_x(T_x = \infty)}.$$

Idées de preuve. Pour tout $k \geq 1$ on calcule $\mathbb{P}_x(N_x \geq k + 1)$ en utilisant la propriété de Markov fort :

$$\mathbb{P}_x(N_x \geq k + 1) = \mathbb{E}_x[\mathbb{E}_{X_{T_x}}(\mathbf{1}_{\{N_x \geq k\}}) \mathbf{1}_{\{T_x < \infty\}}] = \mathbb{P}_x(N_x \geq k) \mathbb{P}_x(T_x < \infty).$$

Comme $\mathbb{P}_x(N_x \geq 1) = 1$ on en déduit par induction

$$\mathbb{P}_x(N_x \geq k) = (\mathbb{P}_x(T_x < \infty))^{k-1}.$$

Si $\mathbb{P}_x(T_x < \infty) = 1$ alors $\mathbb{P}_x(N_x = \infty) = 1$. Sinon,

$$\mathbb{E}_x[N_x] = \sum_{k \geq 1} \mathbb{P}_x(N_x \geq k) = \frac{1}{1 - \mathbb{P}_x(T_x < \infty)}$$

□

2.4.2. *Propriétés de classe.*

Proposition 2.14. Soit $x, y \in E$. On suppose que x est récurrent et que $x \rightarrow y$.

- (1) Alors y est récurrent. Par suite, les éléments d'une même classe sont tous récurrents, ou tous transients. On parle de classe récurrente ou classe transiente.
- (2) De plus $y \rightarrow x$. Par suite, si une classe n'est pas fermée, elle est nécessairement transiente. Enfin,

$$\mathbb{P}_x(T_y < \infty) = \mathbb{P}_y(T_x < \infty) = 1.$$

- (3) (Corollaire) Une classe fermée et de cardinal fini est récurrente.

Cette proposition permet de déterminer la nature de tous les états d'une chaîne de Markov sur un espace d'états E fini : les classes fermées sont récurrentes, les autres transientes. En particulier :

Corollaire 2.15. Une chaîne de Markov irréductible sur un espace fini est récurrente.

Remarque 2.16. Supposons la chaîne de Markov irréductible sur E (de cardinal fini ou infini).

Dans le cas récurrent, tous les points de E sont visités infiniment souvent : pour $x, y \in E$,

$$\mathbb{P}_x(X_n = y \text{ pour une infinité de } n) = 1$$

Dans le cas transient, tous les sous-ensembles finis de E sont visités un nombre fini de fois : pour $A \subset E$ de cardinal fini,

$$\mathbb{P}_x(X_n \in A \text{ pour une infinité de } n) = 0.$$

Il est alors parfois utile, pour mener à bien les calculs, de considérer la *fonction de Green* associée à la chaîne de Markov, définie de la manière suivante :

Définition 2.17. Pour $x, y \in E$ on définit

$$G(x, y) = \mathbb{E}_x[N_y] = \mathbb{E}_x \left[\sum_{n \geq 0} \mathbf{1}_{\{X_n = y\}} \right]$$

Proposition 2.18. Pour tout $x \neq y$ dans E

$$G(x, x) = \frac{1}{\mathbb{P}_x(T_x = \infty)}, \quad G(x, y) = \mathbb{P}_x(T_y < \infty)G(y, y)$$

En particulier : x est récurrent si et seulement si $G(x, x) = +\infty$. De plus, on a toujours : pour tout $x \neq y$, $G(x, y) \leq G(y, y)$. Enfin $x \rightarrow y$ si et seulement si $G(x, y) > 0$.

2.4.3. *Instants de retours.* Soit $x \in E$ et (X_n^x) une chaîne de Markov sur E de probabilité de transition P partant de $X_0 = x$. On introduit la suite des instants successifs de retours en x , par récurrence :

$$T_x^1 = \inf \{k > 0 ; X_k^x = x\}, \quad T_x^{n+1} = \inf \{k > T_x^n ; X_k^x = x\}$$

avec la convention $\inf \emptyset = \infty$. Le nombre de visites en x avant le temps n est noté N_x^n et est défini par

$$N_x^n = \sum_{k=0}^n \mathbf{1}_{\{X_k^x = x\}}.$$

Ces nombres de visites sont reliés aux temps de passage par les relations suivantes :

$$N_x^n \geq k + 1 \Leftrightarrow T_x^k \leq n.$$

La propriété de Markov fort nous permet également de montrer :

Proposition 2.19. Soit $x \in E$. Sur l'évènement $T_x^n < \infty$, les v.a. $T_x, T_x^2 - T_x^1, \dots, T_x^n - T_x^{n-1}$ sont i.i.d.

2.5. Mesures invariantes. À partir de maintenant on supposera la chaîne de Markov (X_n) irréductible. Dans le cas général, on pourra néanmoins appliquer les résultats suivantes à la chaîne de Markov restreinte à une classe fermée.

Les questions de récurrence ou transience s'intéressent plutôt au comportement asymptotique de la chaîne (les états sont-ils visités infiniment souvent ?). On va maintenant regarder des aspects plus quantitatifs, par exemple :

- combien de temps (X_n) passe-t-elle dans les différents états (en proportion) ?
- combien de temps (X_n) met-elle pour revenir à son point de départ (en espérance, i.e. calculer $\mathbb{E}_x[T_x]$) ?
- la loi de (X_n) admet-elle une limite quand n est grand ?

Pour tout n , on note μ_n la loi de (X_n) sur E (que l'on peut considérer comme un vecteur ligne, possiblement infini). On rappelle : pour tout n , $\mu_{n+1} = \mu_n P$. Ainsi, si $\mu_n \rightarrow \mu$ on a de bonnes raisons de penser que μ devrait vérifier $\mu = \mu P$.

Définition 2.20. Une mesure μ sur E (non nécessairement de masse finie) est dite invariante pour P si $\mu = \mu P$, i.e.: pour tout $y \in E$

$$\mu(y) = \sum_{x \in E} \mu(x) P(x, y).$$

Si $\mu(E) = 1$, alors μ est une loi de probabilité invariante.

On s'intéresse maintenant à l'existence de telles mesures. La première façon d'en trouver est de rechercher parmi une certaine famille de mesures :

Définition 2.21. Une mesure μ sur E est réversible pour P si, pour tout $x, y \in E$

$$\mu(x) P(x, y) = \mu(y) P(y, x)$$

Une mesure réversible est invariante, mais l'inverse n'est pas vrai. Ces mesures sont souvent plus faciles à trouver s'il en existe.

2.5.1. Cas E fini. Dans ce cas, la relation $\mu = \mu P$ signifie que μ est un vecteur propre à gauche de P pour la valeur propre 1 (ou encore que ${}^T \mu$ est un vecteur propre à droite de ${}^T P$). On sait déjà que 1 est une valeur propre car la somme des lignes vaut 1, donc on sait qu'il existe $\mu \neq 0$ tel que $\mu P = \mu$. Il faut donc comprendre si l'on peut choisir μ à coefficients positifs. C'est la conséquence (par exemple) d'un résultat d'algèbre (le théorème de Perron-Frobenius) mais il peut aussi être démontré par un argument probabiliste (cf. cas général).

Proposition 2.22. Si E est de cardinal fini, alors il existe une mesure invariante μ pour P . De plus :

- Si la chaîne est irréductible alors il existe une unique loi de probabilité invariante π et elle vérifie $\pi(x) > 0$ pour tout $x \in E$.
- Si la chaîne n'est pas irréductible, toute mesure invariante pour P s'écrit comme combinaison convexe des mesures invariantes des chaînes restreintes à chaque composante récurrente.

2.5.2. Cas général. Dans ce cas il n'existe pas toujours de loi de probabilité invariante.

Proposition 2.23. Supposons la chaîne (X_n) récurrente irréductible. Alors, il existe une mesure invariante (non nulle), unique à un facteur près.

Soit $x \in E$. La mesure invariante notée μ_x est donnée par

$$\mu_x(y) = \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \mathbf{1}_{\{X_n=y\}} \right]$$

et c'est l'unique mesure invariante qui vérifie $\mu_x(x) = 1$.

Dans le cas transient, on peut juste dire :

Proposition 2.24. *Si π est une loi de probabilité invariante et si $x \in E$ est transient, alors $\pi(x) = 0$. En particulier, si la chaîne est transiente et irréductible, il ne peut pas exister de loi de probabilité invariante non nulle.*

2.6. Réurrence positive. Grâce à la formule explicite des mesures invariante μ_x dans le cas récurrent, on peut donner un critère d'existence de loi de probabilité invariante. Notons que sa masse totale est égale à

$$\mu_x(E) = \sum_{y \in E} \mu_x(y) = \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \sum_{y \in E} \mathbf{1}_{\{X_n=y\}} \right] = \mathbb{E}_x[T_x]$$

Ceci nous mène au résultat suivant :

Proposition 2.25. *Supposons la chaîne de Markov (X_n) irréductible récurrente. Alors il n'existe que deux situations possibles :*

- (1) *Toutes les mesures invariantes ont une masse infinie. Dans ce cas : pour tout $x \in E$, $\mathbb{E}_x[T_x] = \infty$. La chaîne est dite récurrente nulle.*
- (2) *Toutes les mesures invariantes ont une masse totale finie. Il existe alors une unique loi de probabilité invariante π , qui vérifie : pour tout $x \in E$,*

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]} > 0,$$

et la chaîne est dite récurrente positive.

On en déduit alors, d'après ce qui précède :

Corollaire 2.26. *Toute chaîne de Markov irréductible sur un espace d'états fini est récurrente positive.*

Dans le cas général, l'ensemble des mesures de probabilité invariantes d'une chaîne est formé des combinaisons convexes des meures invariantes des chaînes restreintes aux classes de communication récurrentes positives.

Exemple 2.27. *Dans l'exemple de la ruine du joueur, les lois de probabilité invariantes sont les combinaisons convexes $p\delta_0 + (1-p)\delta_K$ pour $p \in [0, 1]$.*

2.7. Théorèmes limites.

Théorème 2.28 (Théorème ergodique). *On suppose la chaîne (X_n) irréductible récurrente.*

- (1) *Si la chaîne est récurrente positive, notons π son unique loi de probabilité invariante. Alors, pour tout $x \in E$, pour toute loi initiale μ sur E ,*

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\mu - p.s.} \pi(x).$$

Plus généralement, pour toute fonction $f : E \rightarrow \mathbb{R}$ intégrable par rapport à π

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\mu - p.s.} \int_E f d\pi = \sum_{x \in E} \pi(x) f(x)$$

(2) Si la chaîne est récurrente nulle, alors, pour tout état $x \in E$, pour toute loi initiale μ sur E ,

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\mu - p.s.} 0.$$

Lorsque l'on prend l'espérance dans les convergences ci-dessus, le théorème de convergence dominée nous permet d'en déduire, dans le cas irréductible récurrent :

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}_\mu(X_k = x) \xrightarrow[n \rightarrow \infty]{} \frac{1}{\mathbb{E}_x[T_x]}.$$

Ce n'est pas suffisant pour en déduire que la suite $(\mathbb{P}_\mu(X_n = x))_{n \geq 0}$ converge vers $\pi(x)$. Un contre-exemple typique est celui de la matrice de transition donnée par

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

qui a pour loi invariante $\pi = (\frac{1}{2}, \frac{1}{2})$ mais pour laquelle $\mathbb{P}_0(X_n = 0) = 0$ dès que n est pair. On a donc besoin d'une notion supplémentaire :

Définition 2.29. La période d'un état $x \in E$ est

$$d(x) := \text{PGCD}\{n \geq 1 ; P^n(x, x) > 0\}$$

Si $d(x) = 1$, x est dit apériodique.

Dans l'exemple précédent, $d(x) = 2$ pour tout état x . Notons que si $P(x, x) > 0$ alors immédiatement $d(x) = 1$.

Proposition 2.30. Si x et y communiquent, alors $d(x) = d(y)$: la période est la même pour tous les états d'une même classe.

Théorème 2.31 (Convergence en loi). On suppose que la chaîne (X_n) est irréductible, récurrente positive, et apériodique. Alors, pour tout $x \in E$ et toute loi initiale μ sur E ,

$$\mathbb{P}_\mu(X_n = x) \xrightarrow[n \rightarrow \infty]{} \pi(x).$$

Si la chaîne est irréductible, récurrente nulle et apériodique, ou transiente, alors

$$\mathbb{P}_\mu(X_n = x) \xrightarrow[n \rightarrow \infty]{} 0.$$

2.8. Quelques exemples.

2.8.1. La marche aléatoire simple symétrique sur \mathbb{Z}^d . La marche aléatoire simple sur \mathbb{Z}^d peut être définie de manière équivalente de la manière suivante : $X_0 \in \mathbb{Z}^d$ aléatoire, et pour tout $n \geq 0$,

$$X_{n+1} = X_n + U_{n+1}$$

où les $(U_i)_{i \geq 1}$ sont des v.a.i.i.d. uniformes dans $\{-e_i, e_i ; i = 1, \dots, d\}$, avec (e_i) base de \mathbb{Z}^d .

Théorème 2.32 (Polya, 1921). Si $d = 1$ ou 2 alors la marche aléatoire est récurrente. Si $d \geq 3$ elle est transiente.

Proof. Le caractère transient ou récurrent d'une chaîne irréductible ne dépend pas de son point de départ. On suppose donc $X_0 = 0$ p.s.

(1) On suppose $d = 1$. Pour n impair, on a $P^n(0, 0) = 0$, tandis que pour $n = 2k$ on a

$$P^{2k}(0, 0) = \mathbb{P}(\text{autant de pas à droite qu'à gauche}) = \binom{2k}{k} \frac{1}{2^{2k}} \underset{k \rightarrow \infty}{\sim} \frac{1}{\sqrt{\pi k}}.$$

On calcule alors

$$\mathbb{E}_0[N_0] = \mathbb{E}_0 \left[\sum_{n \geq 0} \mathbf{1}_{\{X_n=0\}} \right] = \sum_{k=0}^{\infty} \mathbb{P}_0(X_{2k} = 0) = \sum_{k=0}^{\infty} \binom{2k}{k} \frac{1}{2^{2k}}$$

qui est une série divergente, donc $\mathbb{E}_0[N_0] = +\infty$ et la chaîne est récurrente.

(2) On suppose $d = 2$. On note $X_n = (X_n^1, X_n^2)$ et $U_n = (U_n^1, U_n^2)$. Les variables

$$S_n = U_n^1 + U_n^2, \quad D_n = U_n^1 - U_n^2$$

sont indépendantes et de même loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. On a alors

$$\begin{aligned} P^{2k}(0, 0) &= \mathbb{P}_0(X_{2k}^1 + X_{2k}^2 = X_{2k}^1 - X_{2k}^2 = 0) \\ &= \mathbb{P}(S_1 + S_2 + \dots + S_{2k} = 0) \mathbb{P}(D_1 + D_2 + \dots + D_{2k} = 0) \sim \frac{1}{\pi k} \end{aligned}$$

d'où la récurrence de la chaîne.

(3) Pour $d \geq 3$, l'intuition nous dit que la chaîne va devenir transiente (la série devient convergente). Pour une preuve rigoureuse (à l'aide des fonctions caractéristiques) est disponible dans [1, Section 2.3.2].

□

2.8.2. *La marche aléatoire simple asymétrique sur \mathbb{Z} .* On suppose maintenant $d = 1$ et les v.a. (U_i) vérifient $\mathbb{P}(U_i = 1) = 1 - \mathbb{P}(U_i = -1) = p \neq \frac{1}{2}$. Dans ce cas, la marche aléatoire est transiente.

En effet, dans ce cas

$$X_n = \sum_{k=1}^n U_k$$

qui est une somme de v.a.i.i.d. et d'après la loi des grands nombres, on a

$$\frac{1}{n} \sum_{k=1}^n U_k \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[U_1] = 2p - 1 \neq 0$$

Ainsi,

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} \begin{cases} +\infty & \text{si } p > \frac{1}{2} \\ -\infty & \text{si } p < \frac{1}{2} \end{cases}$$

et la chaîne est transiente.

2.8.3. *Probabilités d'absorption.* Dans le cas où la chaîne de Markov n'est pas irréductible, et où il y a plusieurs classes fermées, on se demande dans quelle classe la chaîne de Markov va se retrouver "bloquée" au bout d'un certain temps. On définit alors :

Définition 2.33. Soit C une classe fermée (pour la chaîne de Markov X). On note τ_C son temps d'atteinte :

$$\tau_C(X) = \inf \{n \geq 1 ; X_n \in C\}.$$

Pour tout $x \in E$, la probabilité d'absorption par C partant de x est donnée par

$$q_C(x) = \mathbb{P}_x(\tau_C(X) < \infty)$$

On a évidemment : $q_C(x) = 1$ si $x \in C$, et aussi $q_C(x) = 0$ si x appartient à une autre classe fermée différente de C .

Si x appartient à une classe non fermée, alors, pour atteindre C depuis X , il faut l'atteindre à partir de la position X_1 . On peut donc décaler la chaîne d'un rang, et utiliser la propriété de Markov :

$$q_C(x) = \mathbb{P}_x(\tau_C(X^{+1}) < \infty) = \sum_{y \in E} \mathbb{P}_x(X_1 = y, \tau_C(X^{+1}) < \infty) = \sum_{y \in E} \mathbb{P}_x(X_1 = y) \mathbb{P}_y(\tau_C(X) < \infty).$$

Ainsi, si on note \mathcal{T} la réunion des classes non fermées, on obtient

$$q_C(x) = \sum_{y \in \mathcal{T}} P(x, y) q_C(y) + \sum_{y \in C} P(x, y).$$

C'est un système linéaire d'équations, d'inconnues $q_C(x)$ pour $x \in \mathcal{T}$. On peut montrer, si \mathcal{T} est fini, qu'il admet une unique solution.

REFERENCES

- [1] M. Benaïm, N. El Karoui, *Promenade aléatoire*, Ed. École Polytechnique. 2007.

Introduction à la statistique

Cours d'agrégation, option A

2012-2013

1 Exemple et problématique

De manière générale, le rôle du statisticien est d'extraire des informations de données mesurées, sans connaître *a priori* le modèle sous-jacent.

Exemple 1.1 *On modélise l'intention de vote des électeurs à un référendum par 0 pour non et 1 pour oui. On cherche à estimer la probabilité $\theta \in [0, 1]$ qu'un électeur vote oui. Pour ce faire, on interroge n personnes choisies au hasard, dont les réponses x_1, \dots, x_n sont considérées comme des réalisations de v. a. X_1, \dots, X_n i.i.d de même loi de Bernoulli $(1 - \theta)\delta_0 + \theta\delta_1$. **Objectif** : estimer le paramètre θ .*

Exemple 1.2 *Une entreprise de téléphonie suppose que la distribution du nombre journalier des appels passés par ses clients est une loi de Poisson de paramètre θ inconnu. Cette hypothèse, appelée H_0 , lui permet de fixer ses tarifs, et elle souhaite donc la tester. Pour cela, elle mesure sur une journée le nombre d'appels téléphoniques émis, et dispose de l'observation de n v.a. X_1, \dots, X_n i.i.d de loi inconnue. **Objectif** : accepter ou rejeter l'hypothèse H_0 , et, en cas d'acceptation, estimer le paramètre θ de la loi.*

La statistique paramétrique consiste à considérer qu'un phénomène aléatoire obéit à une loi de probabilité, qui appartient à une famille de loi indexée par un paramètre. Le choix de la famille constitue la *modélisation*. La véritable valeur du paramètre est inconnue, et l'on cherche à la déterminer à l'aide d'un *échantillon*. On n'obtiendra jamais une valeur exacte pour ce paramètre inconnu, mais un *intervalle de confiance*, c'est-à-dire un sous-ensemble de l'ensemble des paramètres qui contient la véritable valeur avec au moins une certaine probabilité p , appelée *niveau de confiance*. On parle de *risque* pour désigner la probabilité $1 - p$, et on cherchera à le minimiser. Pour estimer le paramètre, on construit une fonction de l'échantillon, appelée *estimateur*, qui converge, en un certain sens, vers la valeur recherchée du paramètre.

Remarque – Un modèle statistique est toujours faux. Ce n'est qu'une approximation simple de la réalité qu'on espère pas trop mauvaise. Le choix d'un modèle doit toujours être justifié et critiqué. Il repose sur la connaissance du phénomène étudié (par exemple, des représentations graphiques).

2 Rudiments de statistique paramétrique

Définition 2.1

ǎ

1. *Un modèle statistique est la donnée d'un espace mesurable et d'un tribu (Ω, \mathcal{A}) ainsi que d'une famille $(\mathbb{P}_\theta)_{\theta \in \Theta}$ de lois de probabilités sur cet espace.*
2. *Lorsque $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$, le modèle est dit paramétrique. Sinon, il est dit non-paramétrique.*

3. Une observation X est une variable aléatoire à valeurs dans Ω dont la loi inconnue μ appartient à $\{\mathbb{P}_\theta ; \theta \in \Theta\}$. On dispose de données x_1, \dots, x_n , qui constituent un échantillon, et que l'on suppose être des réalisations $X_1(\omega), \dots, X_n(\omega)$ de v.a. i.i.d de même loi μ .

Remarque – Si le modèle est paramétré par θ , on peut espérer dire des choses sur θ à condition que $\theta \rightarrow \mathbb{P}_\theta$ soit injective. On dit alors que le modèle est *identifiable*.

Exemple 1.1 : $\Omega = \{0, 1\}$, \mathcal{A} est l'ensemble des parties de Ω , $\mathbb{P}_\theta = (1 - \theta)\delta_0 + \theta\delta_1$, et $\Theta =]0, 1[$.

Exemple 1.2 : $\Omega = \mathbb{N}$, \mathcal{A} est l'ensemble des parties de Ω ,
 $\Theta = \{\text{fonctions de répartition nulle sur } \mathbb{R}_- \text{ et constante sur tous les intervalles } [k, k + 1]\}$,
 $\mathbb{P}_\theta = \text{loi associée à la fonction de répartition } \theta$.

Dans toute la suite, on suppose donné un modèle statistique identifiable que l'on note $(\Omega, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$.

3 Méthodes d'estimation

L'un des objectifs de la statistique paramétrique est de déterminer au mieux le paramètre θ de la loi inconnue en utilisant les données.

Définition 3.1 Soit (X_1, \dots, X_n) un échantillon de loi \mathbb{P}_θ et $g : \Theta \rightarrow \mathbb{R}^d$ une fonction borélienne. Un estimateur G_n de $g(\theta)$ est une fonction borélienne de X_1, \dots, X_N à valeurs dans $g(\Theta)$, indépendante de θ .

Un estimateur peut avoir différentes propriétés qui illustrent son bon comportement dans l'estimation de $g(\theta)$. On note \mathbb{E}_θ et Var_θ l'espérance et la variance sous la loi \mathbb{P}_θ .

Définition 3.2 On dit que l'estimateur G_n de $g(\theta)$ est

1. sans biais si

$$\forall \theta \in \Theta, \mathbb{E}_\theta[G_n] = g(\theta)$$

La quantité $g(\theta) - \mathbb{E}_\theta[G_n]$ est appelée le biais.

2. asymptotiquement sans biais si

$$\forall \theta \in \Theta, \lim_{n \rightarrow \infty} \mathbb{E}_\theta[G_n] = g(\theta)$$

3. consistant (ou convergent) si

$$\forall \theta \in \Theta, G_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\theta} g(\theta) \quad (\text{convergence en probabilité})$$

4. fortement consistant (ou fortement convergent) si

$$\forall \theta \in \Theta, G_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\theta\text{-p.s.}} g(\theta) \quad (\text{convergence presque sûre})$$

Exemple 1.1 : On estime θ par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

\bar{X}_n est sans biais, et fortement consistant d'après la loi des grands nombres.

3.1 Méthode des moments

On suppose donné un échantillon X_1, \dots, X_n de loi \mathbb{P}_θ . On veut estimer (en supposant que la quantité soit bien définie)

$$g(\theta) = \mathbb{E}_\theta[\phi(X_1)] = \int \phi(x) d\mathbb{P}_\theta(x)$$

On utilise pour cela la convergence presque sûre :

$$\frac{\phi(X_1) + \dots + \phi(X_n)}{n} \xrightarrow[n \rightarrow \infty]{} g(\theta)$$

Par exemple, on estime $\mathbb{E}_\theta[X_1]$ par \bar{X}_n , et $\text{Var}_\theta[X_1]$ par

$$\frac{X_1^2 + \dots + X_n^2}{n} - \left(\frac{X_1 + \dots + X_n}{n} \right)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Plus généralement, la méthode des moments consiste à estimer $g(\theta) = \Phi(\mathbb{E}_\theta[\phi_1(X_1)], \dots, \mathbb{E}_\theta[\phi_p(X_1)])$ par

$$\Phi \left(\frac{1}{n} \sum_{i=1}^n \phi_1(X_i), \dots, \frac{1}{n} \sum_{i=1}^n \phi_p(X_i) \right)$$

Avantage de cette méthode : il est souvent très facile d'exhiber des estimateurs simples, sans de grosses difficultés de calculs.

Inconvénient de cette méthode : les estimateurs obtenus peuvent avoir de mauvaises performances.

3.2 Maximum de vraisemblance

On suppose donné un échantillon X_1, \dots, X_n de loi \mathbb{P}_θ . On supposera ici que \mathbb{P}_θ admet une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d , notée f_θ , ou par rapport à la mesure de comptage sur un ensemble dénombrable.

Définition 3.3 La vraisemblance de l'échantillon (X_1, \dots, X_n) est la fonction définie sur Θ par

$$V_{X_1, \dots, X_n}(\theta) = \prod_{i=1}^n f_\theta(X_i)$$

On note $\hat{\theta}(X_1, \dots, X_n)$ le point qui maximise la fonction $V_{X_1, \dots, X_n}(\theta)$. Cette fonction, à valeurs dans Θ est appelée estimateur du maximum de vraisemblance de θ .

Remarque – Cet estimateur est naturel puisqu'il conduit à privilégier la valeur de θ la "plus probable" au vu de l'observation. Plutôt que de maximiser la vraisemblance, on peut de manière équivalente maximiser le logarithme de la vraisemblance (la log-vraisemblance).

Avantages de cet estimateur :

1. Il peut être construit dans de nombreux cas.
2. On sait montrer les bonnes propriétés de cet estimateur dans des cas très généraux (familles exponentielles par exemple).

Exemple 1.1 : La vraisemblance pour ce modèle est donnée par :

$$V_{X_1, \dots, X_n}(\theta) = \theta^{X_1 + \dots + X_n} (1 - \theta)^{n - (X_1 + \dots + X_n)}$$

Le maximum est atteint en un unique point, \bar{X}_n .

Exemple 1.2 : L'application de cette méthode n'est pas possible si on ne connaît pas la distribution du nombre journalier d'appels. Si la distribution est une loi de Poisson, la vraisemblance s'écrit :

$$V_{X_1, \dots, X_n}(\theta) = e^{-n\theta} \theta^{X_1 + \dots + X_n} \prod_{i=1}^n \frac{1}{X_i!}$$

Le maximum est atteint en un unique point, \bar{X}_n .

3.3 Autres propriétés des estimateurs

Définition 3.4 Le risque quadratique de l'estimateur G_n sous \mathbb{P}_θ est

$$\mathcal{R}(\theta, G_n) = \mathbb{E}_\theta [\|G_n - g(\theta)\|^2]$$

où $\|\cdot\|$ désigne la norme euclidienne.

Définition 3.5 Soit G'_n un autre estimateur de $g(\theta)$. On dit que G_n est préférable à G'_n si, pour tout $\theta \in \Theta$,

$$\mathcal{R}(\theta, G_n) \leq \mathcal{R}(\theta, G'_n)$$

Exemple 1.1 L'estimateur $G_n = \bar{X}_n$ de θ est préférable à l'estimateur $G'_n = X_1$.

Définition 3.6 On dit que l'estimateur G_n est asymptotiquement normal si il existe une suite normalisatrice $(v_n)_{n \in \mathbb{N}^*}$ qui croît vers l'infini telle que, pour tout $\theta \in \Theta$,

$$v_n(G_n - g(\theta)) \xrightarrow[n \rightarrow \infty]{\text{loi sous } \mathbb{P}_\theta} \mathcal{N}(0, 1)$$

Exemple 1.1 D'après le théorème central limite, l'estimateur \bar{X}_n est asymptotiquement normal, avec comme suite normalisante $v_n = \frac{\sqrt{n}}{\theta(1-\theta)}$.

4 Régions de confiance

4.1 Intervalles de confiance

On voudrait construire, à l'aide d'un échantillon, un intervalle dans lequel on espère que se trouve le paramètre recherché θ , ou plus précisément : on contrôlera la probabilité que θ appartienne à cet intervalle. Pour quantifier la qualité d'un intervalle de confiance I , on choisit un *niveau de confiance* $1 - \alpha$ proche de 1, tel que $\theta \in I$ avec probabilité $1 - \alpha$.

Définition 4.1 Soit $\alpha \in [0, 1]$. On cherche à estimer $g(\theta)$ pour $\theta \in \Theta$ à l'aide d'une observation X_1, \dots, X_n .

Une région de confiance de $g(\theta)$ de niveau de confiance $1 - \alpha$ est un ensemble (dépendant de l'observation) $I_n(X_1, \dots, X_n) \subset g(\Theta)$, mesurable, tel que

$$\forall \theta \in \Theta, \mathbb{P}_\theta[g(\theta) \in I_n] = 1 - \alpha$$

On parle d'intervalle de confiance par excès lorsque le signe d'égalité est remplacé par \geq .

Remarque – Les valeurs usuelles de α sont 0.01, 0.05, 0.1. A niveau de confiance fixé, un intervalle de confiance est d'autant meilleur qu'il est de "taille petite".

Remarque – Un intervalle de confiance n'est pas unique. Si les lois \mathbb{P}_θ sont symétriques, un choix raisonnable est de choisir un intervalle symétrique.

On donne maintenant quelques méthodes pour construire des intervalles de confiance.

4.1.1 Application de l'inégalité de Bienaymé-Tchebichev

Rappelons que si X est une variable aléatoire de carré intégrable, alors

$$\forall \varepsilon > 0, \mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

Appliquons cette inégalité à l'exemple 1.1, où on estime θ à l'aide de \bar{X}_n :

$$\forall \varepsilon > 0, \mathbb{P}_\theta[|\bar{X}_n - \theta| \geq \varepsilon] \leq \frac{\theta(1-\theta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

On obtient ainsi un intervalle de confiance par excès de niveau $1 - \alpha$ en considérant

$$I_n = \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}} \right]$$

Remarque – La majoration obtenue par l'application de l'inégalité de Bienaymé-Tchebichev n'est pas très précise, en particulier si la vraie valeur du paramètre est loin de 1/2.

4.1.2 Méthodes des quantiles

Définition 4.2 Soit F la fonction de répartition d'une variable aléatoire. Le quantile q_r d'ordre $r \in]0, 1[$ de la loi F est défini par

$$q_r = \inf\{x \in \mathbb{R} ; F(x) \geq r\}$$

Si F est continue, alors $F(q_r) = r$. Si de plus F est strictement croissante, alors q_r est l'unique solution de l'équation $F(x) = r$.

Reprenons l'exemple 1.1. Sous la loi \mathbb{P}_θ , on sait que $n\bar{X}_n$ suit la loi binomiale de paramètre (n, θ) , qu'on note $\mathcal{B}(n, \theta)$. Notons $q_{n,\theta}^-$ et $q_{n,\theta}^+$ les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de la loi $\mathcal{B}(n, \theta)$. Alors, pour tout $\theta \in]0, 1[$,

$$\mathbb{P}_\theta[n\bar{X}_n \in [q_{n,\theta}^-, q_{n,\theta}^+]] \geq 1 - \alpha$$

(Le fait qu'on n'ait seulement une inégalité provient du fait que les variables aléatoires sont discrètes)

Il faut garder à l'esprit que θ est inconnu, il faut donc inverser les quantiles par rapport au paramètre θ . Les fonctions $\theta \rightarrow q_{n,\theta}^-$ et $\theta \rightarrow q_{n,\theta}^+$ étant croissantes (comparer les fonctions de répartition de $\mathcal{B}(n, \theta_1)$

et de $\mathcal{B}(n, \theta_2)$), on peut définir les "inverses" I_n^- et I_n^+ de $\frac{q_{n,\theta}^-}{n}$ et $\frac{q_{n,\theta}^+}{n}$ par

$$I_n^-(x) = \inf \left\{ \theta \in]0, 1[; \frac{q_{n,\theta}^-}{n} \geq x \right\}$$

$$I_n^+(x) = \inf \left\{ \theta \in]0, 1[; \frac{q_{n,\theta}^+}{n} \geq x \right\}$$

Une écriture équivalente est

$$x \geq \frac{q_{n,\theta}^-}{n} \Leftrightarrow \theta \leq I_n^-(x)$$

$$x \leq \frac{q_{n,\theta}^+}{n} \Leftrightarrow \theta \geq I_n^+(x)$$

Dès lors, on a pour tout $\theta \in]0, 1[$,

$$\mathbb{P}_\theta [\theta \in [I_n^+(\bar{X}_n); I_n^-(\bar{X}_n)]] \geq 1 - \alpha$$

On obtient donc un intervalle de confiance par excès pour θ de niveau $1 - \alpha$ en considérant

$$I_n = [I_n^+(\bar{X}_n); I_n^-(\bar{X}_n)]$$

Les bornes de l'intervalle s'obtiennent numériquement via le calcul des quantiles de la loi $\mathcal{B}(n, \theta)$.

4.1.3 Application des grandes déviations et de l'inégalité de Hoeffding

Lemme 4.1 (Inégalité de Hoeffding) Soit (Y_1, \dots, Y_n) une suite de variables indépendantes centrées et, pour tout i , $a_i \leq Y_i \leq b_i$ p.s., alors

$$\forall \lambda > 0, \mathbb{P} \left[\sum_{i=1}^n Y_i \geq \lambda \right] \leq \exp \left(\frac{-2\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2} \right)$$

Pour une preuve, aller voir dans les développements d'agreg.

Appliquons cette inégalité à l'exemple 1.1. On a :

$$\forall \varepsilon > 0, \mathbb{P}_\theta [|\bar{X}_n - \theta| \geq \varepsilon] \leq 2 \exp(-2n\varepsilon^2)$$

On obtient un intervalle de confiance de niveau $1 - \alpha$ en considérant

$$\left[\bar{X}_n - \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)}; \bar{X}_n + \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)} \right]$$

4.2 Intervalles de confiance asymptotiques

Lorsque nous disposons de suffisamment de données X_1, \dots, X_n , et pour les modèles plus classiques, il est souvent plus simple de remplacer la loi de la statistique par son asymptotique lorsque n est grand. Le théorème central limite s'avère alors un excellent outil, et on parle d'*intervalle de confiance asymptotique*.

Définition 4.3 On dit que la suite $(I_n)_{n \geq 1}$ est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour $g(\theta)$ si, pour tout $\theta \in \Theta$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta [g(\theta) \in I_n(X_1, \dots, X_n)] = 1 - \alpha$$

Notons $v = \theta(1 - \theta)$ la variance de X_1 . Le théorème central limite nous dit que

$$\mathcal{L} \left(\sqrt{\frac{n}{v}} (\bar{X}_n - \theta) \right) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, 1) \quad \text{sous } \mathbb{P}_\theta$$

Ainsi, si $q_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$,

$$\mathbb{P}_\theta \left[\theta \in \left[\bar{X}_n - q_{1-\alpha/2} \sqrt{\frac{v}{n}}; \bar{X}_n + q_{1-\alpha/2} \sqrt{\frac{v}{n}} \right] \right] \xrightarrow[n \rightarrow \infty]{} 1 - \alpha$$

Le défaut de ce raisonnement est que l'on ne connaît pas v puisqu'il dépend du paramètre. Pour corriger ceci, on utilise des outils plus élaborés. Le lemme de Slutsky permet de surmonter certaines difficultés. La méthode delta permet d'obtenir des régions de confiance plus intéressantes. Elle permet surtout d'obtenir la normalité asymptotique de nombreux estimateurs pour l'estimation de $g(\theta)$.

4.2.1 Lemme de Slutsky

Lemme 4.2 (Lemme de Slutsky) Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de vecteurs de \mathbb{R}^d et \mathbb{R}^p , tels que

- $X_n \xrightarrow[n \rightarrow \infty]{} X$ en loi, où X est un vecteur aléatoire,
- $Y_n \xrightarrow[n \rightarrow \infty]{} c$ en probabilité, où c est un vecteur déterministe.

Alors, $(X_n, Y_n) \xrightarrow[n \rightarrow \infty]{} (X, c)$ en loi.

Pour une preuve, cf le poly de Vincent Rivoirard.

En reprenant l'exemple 1.1, et puisque $\bar{X}_n(1 - \bar{X}_n) \xrightarrow[n \rightarrow \infty]{} \theta(1 - \theta)$, on obtient :

$$\mathcal{L} \left(\sqrt{\frac{n}{\bar{X}_n(1 - \bar{X}_n)}} (\bar{X}_n - \theta) \right) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, 1) \quad \text{sous } \mathbb{P}_\theta$$

d'où :

$$\mathbb{P}_\theta \left[\theta \in \left[\bar{X}_n - q_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}}; \bar{X}_n + q_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right] \right] \xrightarrow[n \rightarrow \infty]{} 1 - \alpha$$

Pour $n\theta(1 - \theta)$ assez grand, on obtient un intervalle satisfaisant. Un problème survient lorsque θ est proche de 0 ou 1 en raison de la mauvaise qualité de l'estimation de $\theta(1 - \theta)$.

Remarque – On aurait aussi pu choisir un autre estimateur consistant de la variance :

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Lequel choisir ? La méthode delta, ci-dessous, va fournir une alternative en offrant un estimateur naturel, et résoudre le problème lorsque θ est proche de 0 ou 1.

4.2.2 La méthode delta

On suppose dans cette partie que $g : \mathbb{R}^d \rightarrow \mathbb{R}^p$. On note indistinctement $\| \cdot \|$ la norme euclidienne dans \mathbb{R}^d ou \mathbb{R}^p .

Théorème 4.1 On se donne une suite $(U_n)_{n \in \mathbb{N}}$ de vecteurs aléatoires de \mathbb{R}^d et une suite déterministe $(a_n)_{n \in \mathbb{N}}$ telle que $a_n \xrightarrow{n \rightarrow \infty} +\infty$. On suppose :

- $a_n(U_n - U) \xrightarrow{n \rightarrow \infty} V$ en loi pour V et U deux vecteurs de \mathbb{R}^d (U est non aléatoire),
- g est une fonction différentiable en U dont la différentielle est notée $Dg(U) \in \mathbb{R}^{p \times d}$

Alors

$$a_n(g(U_n) - g(U)) \xrightarrow{n \rightarrow \infty} Dg(U) \cdot V \quad \text{en loi}$$

Pour une preuve, cf le poly de Vincent Rivoirard

Corollaire 4.1 On se donne (X_1, \dots, X_n) un échantillon de vecteurs de \mathbb{R}^d . On suppose que X_1 est de carré intégrable et on note μ la moyenne de X_1 et Σ sa matrice de covariance. Si g est une fonction différentiable en U , alors

$$\sqrt{n} (g(\bar{X}_n) - g(\mu)) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, Dg(U)\Sigma Dg(U)^*) \quad \text{en loi}$$

Exemple 1.1 : On obtient les fluctuations de la variance empirique pour $\theta \neq \{0, 1/2, 1\}$:

$$\mathcal{L} \left(\sqrt{n} \left(\sqrt{\bar{X}_n(1 - \bar{X}_n)} - \sqrt{\theta(1 - \theta)} \right) \right) \xrightarrow{n \rightarrow \infty} \mathcal{N} \left(0, \frac{1}{4}(1 - 2\theta)^2 \right)$$

Ceci explique comment de grandes fluctuations apparaissent lorsque θ est proche de 0 ou 1.

Si on réécrit la première convergence en loi, on a

$$\mathcal{L} \left(\sqrt{n} (\bar{X}_n - \theta) \right) \xrightarrow{n \rightarrow \infty} \sqrt{v} \mathcal{N}(0, 1) \quad \text{sous } \mathbb{P}_\theta$$

Il est possible de trouver une fonction g telle que, en appliquant la méthode delta, on puisse supprimer v dans le membre de droite. En effet, prenons $g(x) = 2 \arcsin \sqrt{x}$. On a alors,

$$\mathcal{L} \left(\sqrt{n} \left(2 \arcsin \sqrt{\bar{X}_n} - 2 \arcsin \sqrt{\theta} \right) \right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

Et un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ est donné par

$$I_n = \left[\sin^2 \left(\arcsin \sqrt{\bar{X}_n} - \frac{1}{2\sqrt{n}} q_{1-\alpha/2} \right); \sin^2 \left(\arcsin \sqrt{\bar{X}_n} + \frac{1}{2\sqrt{n}} q_{1-\alpha/2} \right) \right]$$

4.3 Application à la méthode de Monte-Carlo

La méthode de Monte-Carlo permet le calcul de valeurs approchées d'intégrales multiples en utilisant des réalisations i.i.d de loi uniforme. Si $(X_n)_{n \geq 1}$ est une suite de v.a. i.i.d de loi uniforme sur $[0, 1]^m$, et si $f : [0, 1]^m \rightarrow \mathbb{R}$ est une fonction mesurable, alors la loi des grands nombres entraîne la convergence presque sûre suivante :

$$\frac{1}{n} (f(X_1) + \dots + f(X_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}[f(X_1)] = \int_{[0,1]^m} f(x) dx$$

Si la variance de $f(X_1)$ est majorable, on peut utiliser les intervalles de confiance pour la moyenne. On voit alors que la convergence a lieu à la vitesse \sqrt{n} . Cette vitesse est assez lente, comparée aux méthodes déterministes, mais on remarquera qu'aucune régularité sur f n'est requise ici. La méthode de Monte-Carlo reste valable lorsque f est à support quelconque, en prenant par exemple des X_i de loi gaussienne.

4.4 Exercices

Exercice 1. Pour un n -échantillon (X_1, \dots, X_n) constitué de v.a. i.i.d de loi μ d'espérance m et de variance σ^2 . On note

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

1. On suppose que σ^2 est connue. Donner un intervalle de confiance asymptotique pour m .
2. Même question lorsque la variance σ^2 est inconnue.
3. On suppose que m est connue. Donner un intervalle de confiance asymptotique pour σ^2 .
4. Même question lorsque la moyenne m est inconnue et que la loi μ est une loi gaussienne.

Exercice 2. On suppose que (X_1, \dots, X_n) et (Y_1, \dots, Y_n) sont deux échantillons indépendants de loi gaussienne de même variance mais pas forcément de même moyenne. On note $d = m_X - m_Y$ la différence entre les moyennes de X et Y . Donner un intervalle de confiance asymptotique pour d .

Exercice 3.

1. Simuler un n -échantillon de variables de Bernoulli de paramètre p .
2. Donner un intervalle de confiance à 95% pour p .
3. Reproduire N fois la simulation précédente et déterminer le nombre de fois où l'intervalle de confiance proposé contient le véritable paramètre p .
4. Reprendre cette étude avec des échantillons gaussiens.

Exercice 4. On reprend l'exemple 1.1.

1. Tracer les fonctions de répartition de lois binomiale pour deux paramètres différents p_1 et p_2 . Montrer que si $p_1 < p_2$, la courbe pour p_1 est au dessus de celle pour p_2 .
2. Écrire une fonction qui, pour $r, p \in]0, 1[$ et $n \geq 0$ donne le quantile d'ordre r de la loi $\mathcal{B}(n, p)$.
3. Tracer un histogramme de $\sqrt{\frac{n}{v}}(\bar{X}_n - \theta)$ pour $n = 5, 10$, et 100 et pour $\theta = 0, 2$. Pour chacune de ses figures, superposer la densité d'une gaussienne standard pour justifier l'utilisation des intervalles de confiance asymptotiques.

5 Tests d'hypothèse

5.1 Généralités

L'une des fonctions des statistiques est de prendre des décisions, à partir d'observations d'un phénomène aléatoire (ou modélisé comme tel). Par exemple, peut-on considérer qu'un médicament donné est plus efficace qu'un autre? Les gènes codant la couleur des yeux et celle des cheveux sont-ils sur les mêmes chromosomes? Étant donné le nombre de composants défectueux présents à la sortie de l'usine, l'industriel doit-il en arrêter la production? Il y a deux points communs à toutes ces questions : on y répond par oui ou non, et le phénomène sous-jacent est aléatoire. Les tests statistiques vont permettre d'apporter une réponse à ces problèmes de décision.

Dans le cadre de notre modèle statistique $(\Omega, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$, on se donne Θ_0 et Θ_1 disjoints tels que $\Theta_0 \cup \Theta_1 = \Theta$. Au vu d'une observation X , on veut déterminer si $\theta \in \Theta_0$ ou non. On note :

- l'hypothèse nulle $H_0 : \theta \in \Theta_0$,
- l'hypothèse alternative $H_1 : \theta \in \Theta_1$.

L'hypothèse H_0 est dite *simple* si Θ_0 est réduit à un singleton, et *composite* sinon. Nous verrons que H_0 et H_1 ne jouent pas des rôles symétriques.

Définition 5.1 On appelle test d'hypothèse H_0 contre l'hypothèse H_1 toute fonction mesurable de l'observation X , notée $\phi(X)$, à valeurs dans $\{0, 1\}$, telle que :

- si $\phi(X) = 0$, alors on accepte H_0 ,
- si $\phi(X) = 1$, alors on rejette H_0 .

Ainsi, $\phi(X)$ s'écrit $\mathbf{1}_{X \in R}$. L'ensemble R est appelé zone de rejet de H_0 .

Définition 5.2

1. On dit qu'on commet une erreur de première espèce lorsque H_0 est vraie, mais est rejetée par le test. L'erreur de première espèce est donc la fonction définie sur Θ_0 par :

$$\theta \rightarrow \mathbb{P}_\theta(\phi(X) = 1)$$

2. On dit qu'on commet une erreur de seconde espèce lorsque H_1 est vraie, mais que H_0 est acceptée. L'erreur de seconde espèce est donc la fonction définie sur Θ_1 par :

$$\theta \rightarrow \mathbb{P}_\theta(\phi(X) = 0)$$

3. On dit que le test est de niveau α si α majore l'erreur de première espèce :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi(X) = 1) \leq \alpha$$

On considère que rejeter H_0 à tort est plus préjudiciable que qu'accepter H_0 à tort. La stratégie consistera à construire le test (et donc la zone de rejet) de façon à minimiser en priorité l'erreur de première espèce. En particulier, l'hypothèse nulle représentera souvent ce que l'on pense être la réalité.

Exemple 5.1 On vient d'acheter un dé à six faces censé fournir des résultats distribués selon la loi uniforme sur $\{1, \dots, 6\}$. On souhaite valider (ou invalider) cette information du fabricant à partir de l'observation de n lancers de dé. On est donc amené à poser les hypothèses :

$$H_0 : \text{le dé est équilibré,} \quad \text{et} \quad H_1 : \text{le dé est pipé}$$

Exemple 5.2 Un industriel souhaite évaluer la qualité de ses produits. Après l'observation de n pièces, il voudrait estimer la proportion de composants défectueux : si cette proportion est trop grande, il sera contraint d'arrêter temporairement sa production le temps de rectifier les machines d'usinage. Une telle éventualité ayant un certain coût, il ne voudra faire ce choix qu'en cas de grande nécessité.

Notre observation est donc une suite de v.a i.i.d de loi de Bernoulli de paramètre p : la valeur 1 représente un composant défectueux. Après avoir déterminé un seuil p_s (au delà duquel la production devra être arrêtée), la question que l'industriel se pose est : p est-il plus grand ou plus petit que p_s ?

Du point de vue de l'industriel, réparer ses machines entraînerait des frais non négligeables. Il ne souhaite pas en arriver là si cela n'est pas nécessaire, souhaitant minimiser le risque de décider $p > p_s$, alors qu'en réalité $p \leq p_s$. Ainsi, on choisit :

$$H_0 : p \leq p_s \quad \text{et} \quad H_1 : p > p_s$$

Si au contraire, nous nous plaçons du point de vue d'un organisme de qualité, l'intérêt serait d'être le plus sûr possible que le seuil n'est pas atteint. On inverserait alors les deux hypothèses.

On choisit souvent le niveau $\alpha = 0.05$. Une fois le niveau fixé, on choisit en second lieu de contrôler la probabilité de rejeter H_0 lorsque cela est nécessaire.

Définition 5.3 On appelle puissance du test la fonction définie sur Θ_1 par :

$$\theta \rightarrow \mathbb{P}_\theta(\phi(X) = 1)$$

Autrement dit, la puissance est la probabilité d'accepter H_1 alors que H_1 est vraie. L'objectif est donc de construire, parmi les tests de niveau α , ceux dont la puissance est maximale. Les deux motivations sont en conflit :

- plus le niveau est faible, plus la zone de rejet est petite, et alors la puissance diminue,
- plus Θ_1 est grand, plus la zone de rejet est grande, ce qui augmente la puissance, mais également le niveau.

Exemple 5.3 *Le test d'hypothèses $H_0 : p = 0.2$ et $H_1 : p \neq 0.2$ est plus puissant que $H_0 : p = 0.2$ et $H_1 : p = 0.4$.*

Toutes ces considérations peuvent nous aider à déterminer H_0 et H_1 . Par exemple, on prendra pour H_0 :

- une hypothèse qu'on pense être vraie,
- une hypothèse de prudence (critère de coût, de sécurité, ...)
- la seule facile à formuler

La démarche est alors la suivante :

1. Choix de H_0 et H_1 .
2. Détermination de la statistique de test $\phi(X)$: on doit en connaître la loi sous H_0 .
3. Étude de l'allure de la zone de rejet en fonction de la forme de H_1 .
4. Calcul de la zone de rejet, à niveau α préalablement fixé.
5. Utilisation des données, calcul, et conclusion.

Définition 5.4 *On dit qu'un test de niveau α est sans biais pour H_0 si sa puissance est toujours plus grande que α :*

$$\forall \theta \in \Theta_1, \mathbb{P}_\theta(\phi(X) = 1) > \alpha$$

Comme précédemment, il peut être plus intéressant d'utiliser les théorèmes limites et de considérer des échantillons de plus en plus grands.

Définition 5.5 *Soit (X_1, \dots, X_n) une observation, et R_n une zone de rejet de niveau α associée aux hypothèses H_0 et H_1 . On dit que le test est de niveau asymptotique α si*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \mathbb{P}_\theta(R_n) \leq \alpha$$

5.2 Test du χ^2

Le test du χ^2 est un exemple relativement simple de test statistique qui permet de tester :

1. l'adéquation à une loi de probabilité sur un ensemble fini (par exemple, est-il raisonnable de penser que les résultats que j'observe sont des réalisations i.i.d d'une loi (p_1, \dots, p_k) sur $\{1, \dots, k\}$?),
2. l'indépendance de deux caractères mesurés sur un même individu,
3. l'homogénéité de plusieurs échantillons (par exemple, deux médicaments ont-ils le même effet sur la population ?).

Considérons un phénomène aléatoire de loi inconnue $P = p_1\delta_1 + \dots + p_k\delta_k$. Des considérations extérieures nous font penser que la loi du phénomène est $Q = q_1\delta_1 + \dots + q_k\delta_k$. Partant de n observations, on désire tester l'hypothèse $H_0 : P = Q$ contre l'hypothèse $H_1 : P \neq Q$.

On note les effectifs des n observations indépendantes par N_1^n, \dots, N_k^n (le nombre N_i^n représente le nombre de fois où on a obtenu i dans l'échantillon de taille n). On désigne par p_i^n la distribution empirique associée définie par :

$$p_i^n = \frac{N_i^n}{n}$$

et on note $P_n := (p_1^n, \dots, p_k^n)$.

On définit la pseudo-distance du χ^2 entre deux distributions A et B discrètes sur $\{1, \dots, k\}$ par :

$$\chi^2(A, B) = \sum_{i=1}^k \frac{(A_i - B_i)^2}{B_i}$$

On montre alors que :

Proposition 5.1

1. La variable aléatoire $\sqrt{n} \left(\frac{p_1^n - p_1}{\sqrt{p_1}}, \dots, \frac{p_k^n - p_k}{\sqrt{p_k}} \right)$ converge en loi vers la loi normale centrée sur \mathbb{R}^k de matrice de covariance $I_k - \sqrt{P}(\sqrt{P})^*$.
2. Sous H_0 , $n \cdot \chi^2(P_n, Q)$ converge en loi vers la loi $\chi^2(k-1)$.
3. Sous H_1 , $n \cdot \chi^2(P_n, Q)$ converge en probabilité vers $+\infty$.

Pour une preuve, cf le poly de F. Malrieux, prépa agreg de Rennes + Vincent Rivoirard.

Le fait que la loi limite ne dépende du paramètre P qu'à travers sa taille k est très important ici. On est ainsi capable de construire une statistique de test : $n \cdot \chi^2(P_n, Q)$, et de choisir une région de \mathbb{R} , telle que la statistique appartienne à cette région avec une probabilité (asymptotique) donnée.

Plus précisément, notons χ_α^2 le quantile d'ordre $1 - \alpha$ de la loi $\chi^2(k-1)$. Un test du χ^2 de niveau asymptotique α consiste en ce qui suit, pour n "assez grand" :

- Si $n \cdot \chi^2(P_n, Q) > \chi_\alpha^2$, alors on choisit H_1 ,
- Si $n \cdot \chi^2(P_n, Q) \leq \chi_\alpha^2$, alors on choisit H_0 .

Avant de donner plusieurs exemples, ajoutons un autre résultat, qui n'est qu'une conséquence directe de la loi des grands nombres mais qui joue un rôle essentiel dans l'élaboration des tests du χ^2 :

Proposition 5.2 *Sous H_1 , $\chi^2(P_n, Q)$ converge presque sûrement vers $\sum_{i=1}^k \frac{(p_i - q_i)^2}{q_i}$.*

Remarque – Les tests du χ^2 ne donnent vraiment d'information que si l'hypothèse H_0 est rejetée. En effet, si H_0 n'est pas rejetée, il se peut bien que ce soit parce que la loi p de l'échantillon est dans H_1 , mais tout près de H_0 . Ceci est encore renforcé quand on est obligé de créer des classes pour des lois continues par exemple (voir le paragraphe suivant) : des tas de lois fourniront les mêmes vecteurs de probabilité sur l'ensemble fini. *Moralité* : on se sert du test du χ^2 (ou de tout autre test non-paramétrique) pour invalider un modèle. Si H_0 est rejetée, alors il faut changer de modèle. Si non, c'est que le modèle (bien que simpliste, approximatif... et vraisemblablement faux) est satisfaisant.

5.2.1 Test d'adéquation

Soit (X_1, \dots, X_n) un n -échantillon réel de loi inconnue μ . On suppose que les X_i prennent les valeurs dans k classes I_1, \dots, I_k constituées de k intervalles ou de k nombres selon que la loi μ est discrète ou continue. On note \hat{f}_i la fréquence empirique associée à la classe i :

$$\hat{f}_i = \frac{1}{n} \sum_{m=1}^n \mathbb{1}_{X_m \in I_i}$$

Soit ν une loi de probabilité connue, de même nature que μ . On doit tester l'hypothèse H_0 : "la loi des X_i est ν " contre H_1 : "la loi des X_i n'est pas ν ".

Les $f_i := \nu(I_i)$ sont calculables a priori, et on a alors sous H_0 :

$$n \cdot \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{f_i} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(k-1)$$

Remarque – Le test du χ^2 s'appuie sur deux résultats asymptotiques (une convergence en loi et une convergence en probabilité). Or, on ne dispose jamais que d'un nombre fini d'observations. Toute la question est de savoir si l'on a le droit de faire comme si la limite en loi était une égalité. En pratique, on dit que pour que le test soit valide, il faut que, pour tout $i = 1, \dots, k$, $n \cdot f_i$ soit supérieur ou égal à 5. Si ce n'est pas le cas, il faut regrouper des classes à faible effectif pour atteindre le seuil exigé.

Exemple 5.4 (*Historique*) Pour tester sa théorie génétique, Mendel croisa des pois tous jaunes et lisses et obtint à la première génération des pois jaunes ou verts et lisses ou ridés. Plus précisément, il obtint 315 pois jaunes et lisses, 108 pois verts et lisses, 101 pois jaunes et ridés, et 32 pois verts et ridés. Est-ce que ces observations confirment ou infirment la théorie mendélienne ? Sous cette approche, la proportion p_0 de chacune des classes précédentes est $p_0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$. On teste donc $H_0 : p = p_0$ contre $H_1 : p \neq p_1$. Quel est le résultat obtenu ?

5.2.2 Test d'homogénéité

Soit (X_1, \dots, X_n) un n -échantillon réel de loi inconnue μ_1 et (Y_1, \dots, Y_m) un m -échantillon réel de loi inconnue μ_2 , les deux échantillons étant indépendants l'un de l'autre. On suppose que les X_i et Y_i prennent leurs valeurs dans k classes (I_1, \dots, I_k) . On note \hat{f}_i^X et \hat{f}_i^Y les fréquences empiriques associées à la classe $i \in \{1, \dots, k\}$. On désire tester l'hypothèse H_0 : " μ_1 et μ_2 sont identiques" contre H_1 : " μ_1 et μ_2 sont différentes".

On définit une sorte de fréquence empirique commune aux deux échantillons en posant :

$$\hat{f}_i := \frac{n \cdot \hat{f}_i^X + m \cdot \hat{f}_i^Y}{n + m}$$

On a alors, sous H_0 :

$$n \cdot \sum_{i=1}^k \frac{(\hat{f}_i - \hat{f}_i^X)^2}{\hat{f}_i} + m \cdot \sum_{i=1}^k \frac{(\hat{f}_i - \hat{f}_i^Y)^2}{\hat{f}_i} \xrightarrow[n, m \rightarrow \infty]{\mathcal{L}} \chi^2(k-1)$$

Exemple 5.5 On cherche à invalider l'affirmation qui consiste à dire que toutes les lessives se valent. On utilise 3 lessives appelées A, B, et C. A la sortie du lavage, on classe les vêtements en 3 catégories : très sale (TS), légèrement sale (LS), et propre (P).

	TS	LS	P
A	30	65	205
B	23	56	121
C	75	125	300

Peut-on dire, au niveau 0.05, que toutes les lessives sont identiques ?

5.2.3 Test d'indépendance

Soit $((X_1, Y_1), \dots, (X_n, Y_n))$ un n -échantillon réel de loi inconnue μ . On note μ_1 et μ_2 les lois marginales de μ et l'on suppose que les (X_i, Y_i) prennent respectivement leurs valeurs sur k classes (I_1, \dots, I_k) et

ℓ classes (J_1, \dots, J_ℓ) . On note $\hat{f}_{i,j}$ la fréquence empirique associée à (I_i, J_j) , définie par :

$$\hat{f}_{i,j} := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k \in I_i \text{ et } Y_k \in J_j}$$

On désire tester l'hypothèse $H_0 : \mu = \mu_1 \otimes \mu_2$ contre $H_1 : \mu \neq \mu_1 \otimes \mu_2$.

Pour cela, on définit $f_{i,j}$ par :

$$f_{i,j} := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k \in I_i} \cdot \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{Y_k \in J_j}$$

On a alors, sous H_0 :

$$n \sum_{i=1}^k \sum_{j=1}^{\ell} \frac{(f_{i,j} - \hat{f}_{i,j})^2}{f_{i,j}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2((k-1)(\ell-1))$$

(Pour une preuve, cf le poly de Rennes)

Exemple 5.6 (Yeux et cheveux). Parmi une population de 124 personnes, on note la couleur de leurs yeux et de leurs cheveux :

	blonds	bruns	roux	noirs
bleus	25	9	7	3
gris	13	17	7	10
marrons	7	13	5	8

Les deux critères sont-ils indépendants au niveau 0.05 ?

5.2.4 Exercices

Exercice 1. Montrer que la puissance du test du χ^2 tend vers 1 lorsque n tend vers l'infini.

Exercice 2. Générer un échantillon de taille 100 de loi uniforme sur $[0, 1]$. Le décomposer en k classes de même longueur et faire un test du χ^2 d'ajustement à la loi uniforme. Recommencer en augmentant le nombre de classes, la taille de l'échantillon, puis en changeant de loi : normale, binomiale, Poisson,...

Exercice 3. Générer un échantillon (X_1, \dots, X_n) de taille $n = 1000$ de loi normale centrée réduite. Considérer $Y_i := X_i^2$, puis tester l'indépendance de X et de Y . Faire de même avec d'autres lois : Bernoulli, exponentielle, uniforme...

Exercice 4. Générer deux échantillons de même loi (par exemple uniforme, normale, exponentielle...) et de tailles différentes, par exemple 1000 et 10000. Tester ensuite l'homogénéité de ces deux échantillons. Faire de même avec deux échantillons de lois distinctes mais de même support.

5.3 Test de Kolmogorov-Smirnov

Soit μ une mesure de probabilité sur \mathbb{R} de fonction de répartition F , et $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. de loi μ .

Définition 5.6

1. La mesure empirique d'un échantillon (X_1, \dots, X_n) est définie par

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

2. La fonction de répartition empirique d'un échantillon (X_1, \dots, X_n) est définie par

$$\forall t \in \mathbb{R}, F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq t}$$

Remarque – La loi μ_n est une mesure aléatoire, de fonction de répartition F_n . La fonction F_n est constante par morceaux.

Une conséquence de la loi des grands nombres est que la mesure empirique μ_n converge étroitement vers la loi μ :

$$\forall f \in \mathcal{C}_b^0(\mathbb{R}), \frac{1}{n} \sum_{i=1}^n f(X_i) = \int_{\mathbb{R}} f d\mu_n \xrightarrow[n \rightarrow \infty]{} \int_{\mathbb{R}} f d\mu$$

Ce résultat de convergence étroite des mesures empiriques peut être affiné en une convergence uniforme des fonctions de répartition, comme le précise le théorème de Glivenko-Cantelli :

Théorème 5.1 (Glivenko-Cantelli) *Presque sûrement, la fonction de répartition empirique converge uniformément vers la fonction de répartition de μ :*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \|F_n - F\| \xrightarrow[n \rightarrow \infty]{p.s.} 0$$

Supposons désormais que F est continue, et donc que μ n'a pas de masses ponctuelles. Alors, on peut préciser la vitesse de convergence dans le théorème précédent :

Théorème 5.2 *Si F est continue, alors*

$$\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_{KS}$$

où D_{KS} est une variable aléatoire dont la loi de probabilité sur \mathbb{R}_+ , appelée distribution de Kolmogorov-Smirnov, est donnée par sa fonction de répartition F_{KS} :

$$\forall t > 0, F_{KS}(t) = 1 + 2 \sum_{k=1}^{+\infty} (-1)^k e^{-2k^2 t^2}$$

5.3.1 Test de Kolmogorov-Smirnov

On voudrait savoir si l'échantillon obtenu X_1, \dots, X_n est bien de loi μ_0 . Il est très important de vérifier que la mesure μ_0 n'a pas d'atomes.

D'après le théorème précédent, on va identifier la loi de $\sqrt{n} \|F_n(x) - F(x)\|_{\infty}$ à celle de Kolmogorov-Smirnov pour n assez grand. On peut utiliser les quantiles à 0.95 et 0.99 pour obtenir un test de niveau 0.05 et 0.01 sur l'adéquation de la loi de l'échantillon μ .

Plus précisément, on teste H_0 : "la fonction de répartition réelle F est F_0 " contre H_1 : " $F \neq F_0$ ". On fixe α le niveau du test. On note $D_n = \sqrt{n} \|F_n - F_0\|_{\infty}$, qui se calcule aisément. Alors :

$$D_n \begin{cases} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_{KS} & \text{sous } H_0 \\ \xrightarrow[n \rightarrow \infty]{p.s.} +\infty & \text{sous } H_1 \text{ car } \|F_n - F_0\|_{\infty} \xrightarrow[n \rightarrow \infty]{p.s.} \|F - F_0\|_{\infty} > 0 \end{cases}$$

On choisit comme zone de rejet $R_{\alpha} = [q_{1-\alpha}; +\infty[$ où $q_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi de Kolmogorov-Smirnov.

5.3.2 Exercices

Exercice 1. Générer un échantillon de taille 100 de loi uniforme sur $[0, 1]$, puis tracer sur le même graphique la fonction de répartition empirique et la fonction de répartition théorique. Faire un test de Kolmogorov-Smirnov sur cet échantillon pour un risque de 0.05. Recommencer 100 fois le test. Combien de fois l'hypothèse H_0 est-elle acceptée? Faire la même chose pour la loi normale centrée réduite et comparer au résultat obtenu avec un test du χ^2 .

Exercice 2.

1. Soient (X_1, \dots, X_n) et (Y_1, \dots, Y_m) deux échantillons de lois continues inconnues. Montrer la convergence en loi suivante :

$$\sqrt{\frac{nm}{n+m}} \sup_{x \in \mathbb{R}} |F_n^{(X)}(x) - F_m^{(Y)}(x)| \xrightarrow[n, m \rightarrow \infty]{\mathcal{L}} D_{KS}$$

2. En déduire un test d'homogénéité de Kolmogorov-Smirnov, et le mettre en place pour deux échantillons de loi uniformes sur $[0, 1]$ et de tailles respectives 100 et 1000.