

Philippe CARMONA

NOTES DU COURS
X22M010

Philippe CARMONA

Laboratoire Jean Leray, UMR 6629, Université de Nantes, 92208, F-44322, Nantes cedex 03,
e-mail: philippe.carmona@math.univ-nantes.fr

**NOTES DU COURS
X22M010**

Philippe CARMONA

Résumé. —

Chaîne de Markov à temps discret

SECTION 1

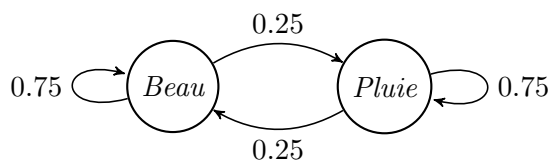
Introduction

Exemple 1.1. — On dispose de statistiques, de tableaux de chiffres, indiquant qu'il fait beau 50% du temps et mauvais 50% du temps. Une première façon naïve de prédire le temps est la suivante

$$\mathbb{P}(\text{beau}) = \mathbb{P}(\text{mauvais}) = 0.5$$

La justification est la loi des grands nombres : l'hypothèse de base est que le climat du lendemain est indépendant du climat des autres jours.

Malheureusement, on constate en examinant les chiffres qu'il y a 3 fois plus de chances que le climat du lendemain reste le même que celui d'aujourd'hui (plutôt qu'il ne change). On représente ce modèle par le diagramme



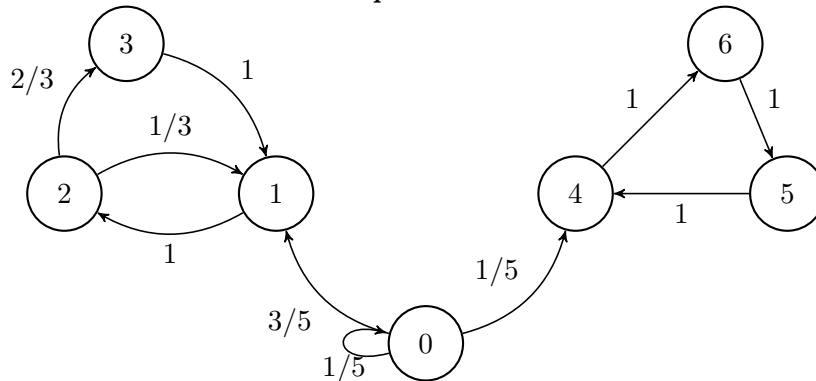
ou encore par la matrice $P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$.

Pour prédire le temps de demain nous avons besoin de

- un mécanisme de transition (décrit par un graphe ou une matrice)
- savoir quel temps il fait aujourd'hui

Considérons un exemple plus évolué

Exemple 1.2. —



On aimerait répondre à la liste de questions suivantes:

- Partant de 0 quelle est la probabilité de toucher 6 ?
- Partant de 1 quelle est la probabilité de toucher 3 ?
- partent de 1 combien de temps en moyenne faut-il pour toucher 3 .
(réponse 3)
- Partant de 1, la proportion de temps passé en 2 est, en temps long, $3/8$
- Partant de 0, la probabilité que je sois en 1 à l'instant n , est pour n grand, proche de $9/32$.

Enfin, étudions un modèle que l'on peut voir comme une chaîne de Markov sur 2 points.

Exemple 1.3. — On suppose que deux machines fonctionnent indépendamment, et identiquement : elles fonctionnent toute la journée (de travail) avec la probabilité p et tombent en panne avec la probabilité $(1 - p)$.

Un technicien, qui ne travaille que la nuit, arrive à réparer, toujours, une machine en panne, mais pas 2. Soit X_n le nombre de machines en panne au n -ième jour.

Un peu de réflexion permet de modéliser X_n comme une chaîne de Markov sur $\{0, 1\}$ de matrice de transition: $\begin{pmatrix} 1 - (1 - p)^2 & (1 - p)^2 \\ p & 1 - p \end{pmatrix}$.

SECTION 2

Définitions et propriétés de base

Soit I un ensemble fini ou dénombrable, appelé espace d'états (nommé aussi S ou E). Une mesure (positive) sur I est une famille $(\lambda_i, i \in I)$ de nombres positifs. C'est une loi de probabilité si $\sum_i \lambda_i = 1$.

Une matrice $P = (p_{ij})$ est stochastique (on dit aussi matrice de transition) si toutes ses lignes sont des probabilités i.e.

$$\forall i, j, p_{ij} \geq 0 \quad \text{et} \quad \forall i, \sum_j p_{ij} = 1$$

Il y a une bijection évidente entre les matrices de transition et les graphes valués donnée par il y a une flèche de i à j valuée par p_{ij} ssi $p_{ij} > 0$.

Définition 1.1. — Une chaîne de Markov homogène à valeurs dans I de matrice de transition P et de loi initiale λ est une famille de variables aléatoires $(X_n, n \in \mathbb{N})$ telle que

1. $\mathbb{P}(X_0 = i) = \lambda(i)$
2. $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n, i_{n+1}}$

On dit alors que $(X_n)_{n \in \mathbb{N}}$ est markov (λ, P) .

Exemple 1.4. — $I = \{1, 2, 3\}$ est l'espace des 'états avec 1 qui signifie en panne, 2 en fonctionnement erratique et en bon fonctionnement. Les transitions sont décrites par la matrice

$$P \begin{pmatrix} 0 & 2 & 0 \\ 0 & 2/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

Question Si je pars d'une machine en bon état, quelle est la probabilité α (resp. β) qu'elle soit en bon état (resp. erratique) dans 4 jours ?

Réponse il y a deux (??) chemins de même probabilité pour obtenir une machine en bon état $\alpha = 2 \frac{1}{2} \frac{1}{3}$

2.1 Caractérisation

Proposition 1.1. — $(X_n)_{0 \leq n \leq N}$ est Markov λ, P ssi pour tous i_0, \dots, i_N

$$\mathbb{P}(X_0 = i_0, \dots, X_N = i_N) = \lambda(i_0) p_{i_0, i_1} \dots p_{i_{N-1}, i_N}$$

Démonstration. — conditionnements successifs et récurrence. □

On note δ_i la masse de Dirac en i et \mathbb{P}_i la loi d'une chaîne (δ_i, P) (on dit que la chaîne est issue de i car $X_0 = i$ ps).

Proposition 1.2. — Si $(X_n)_{n \geq 0}$ est Markov (λ, P) , alors conditionnellement à $X_m = i$ le processus $(X_{m+n}, n \geq 0)$ est Markov (δ_i, P) et est indépendant de (X_0, \dots, X_m) .

Démonstration. — On admet qu'il suffit de montrer que pour tout évènement A de la forme $\{X_0 = i_0, \dots, X_m = i_m\}$ $A \in \mathcal{F}_m^X := \sigma(X_0, \dots, X_m)$ et tout évènement $B = \{X_m = i_m, \dots, X_{m+n} = i_{m+n}\}$ de $\sigma(X_p, p \geq m)$ on a

$$(1) \quad \mathbb{P}(A \cap B \mid X_m = i) = \mathbb{P}(A \mid X_m = i) \mathbb{P}_i(X_0 = i_m, \dots, X_n = i_{m+n})$$

□

Remarque. — En appliquant la proposition précédente à $m = 0$ on obtient que quelle que soit la loi initiale λ de X_0 , du moment qu'elle charge le point i , ($\lambda(i) > 0$), on a pour tout évènement A , $\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i)$. En bref, conditionner la chaîne à valoir i en 0 revient à la faire partir de i .

Proposition 1.3. — On suppose que $(\xi_n)_{n \geq 1}$ est IID à valeurs dans E et indépendante de la variable aléatoire X_0 , et que la suite X_n est définie par la relation

$$X_{n+1} = f(X_n, \xi_{n+1}) \quad (n \geq 1),$$

pour une fonction $f : E \times I \rightarrow I$. Alors $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de probabilités de transition:

$$p_{ij} = \mathbb{P}(f(i, \xi_1) = j)$$

Démonstration. — On vérifie que la matrice P définie ci dessus est bien une matrice de transition.

Ensuite, il suffit, conformément à la proposition ??, de calculer, en utilisant l'indépendance des ξ_k et de X_0 ,

$$\begin{aligned} \mathbb{P}(X_0 = i_0, \dots, X_N = i_N) &= \mathbb{P}(X_0 = i_0, f(i_0, \xi_1) = i_1, f(i_1, \xi_2) = i_2, \dots, f(i_{N-1}, \xi_N) = i_N) \\ &= \mathbb{P}(X_0 = i_0) \prod_{k=0}^{N-1} \mathbb{P}(f(i_k, \xi_{k+1}) = i_{k+1}). \end{aligned}$$

□

2.2 Calculs algébriques On note une mesure positive λ comme un vecteur ligne et donc λP est la mesure $(\lambda P)(j) = \sum_i \lambda(i) p_{i,j}$ On note $p_{i,j}^{(n)}$ l'élément i, j de la matrice P^n et donc on a $p_{i,j}^{(2)} = \sum_k p_{i,k} p_{k,j}$ et en général

$$p_{i,j}^{(n)} = \sum_{i_0=i, i_1, \dots, i_n=j} p_{i_0, i_1} \cdots p_{i_{n-1}, i_n}$$

Proposition 1.4. — Soit $(X_n)_{n \geq 0}$ Markov (λ, P) . Alors pour tout n la loi de X_n est λP^n i.e.

$$\mathbb{P}(X_n = j) = (\lambda P^n)(j)$$

En particulier, pour $\lambda = \delta_i$ et tout $m \geq 0$, on obtient

$$\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{m+n} = j \mid X_m = i) = p_{i,j}^{(n)}$$

Démonstration. — Récurrence sur n , propriété de Markov et formule des probabilités totales:

$$\mathbb{P}(X_{n+1} = j) = \sum_i \mathbb{P}(X_{n+1} = j \mid X_n = i) \mathbb{P}(X_n = i) = \sum_i p_{ij} (\lambda P^n)(i) = (\lambda P^n P)(j).$$

□

Exemple 1.5. — Pour le climat à Nantes, s'il fait beau aujourd'hui, la probabilité qu'il fasse beau dans trois jours est

$$\mathbb{P}(X_3 = 1 \mid X_0 = 1) = p_{1,1}^{(3)} = 9/16$$

car on calcule

$$P^2 = \begin{pmatrix} 10/16 & 6/16 \\ 6/16 & 10/16 \end{pmatrix}$$

SECTION 3

Classification des états

3.1 Structure de classe

Définition 1.2. — On dit que i mène à j et on le note $i \rightarrow j$ si

$$\mathbb{P}_i(\exists n \geq 0 : X_n = j) > 0.$$

On dit que i et j communiquent et on le note $i \leftrightarrow j$ si $i \rightarrow j$ et $j \rightarrow i$.

Proposition 1.5. — Sont équivalentes

- (a) $i \rightarrow j$
- (b) $\exists i_0 = i, i_1, \dots, i_n = j$ tel que $p_{i_0, i_1} \cdots p_{i_{n-1}, i_n} > 0$
- (c) Il existe un chemin dans le graphe de transition menant de i à j .
- (d) $\exists n \geq 0$ tel que $p_{i,j}^{(n)} > 0$.

Démonstration. — On admet les chemins de longueur nulle. (b) et (c) sont identiques; Comme une somme de termes positifs est > 0 ssi il existe un term

> 0 , on en déduit que (b) \iff (d). Enfin l'équivalence entre (a) et (d) découle de la double inégalité

$$p_{i,j}^{(n)} = \mathbb{P}_i(X_n = j) \leq \mathbb{P}_i(\exists m, X_m = j) \leq \sum_m \mathbb{P}_i(X_m = j)$$

En effet, si (a), alors $\mathbb{P}_i(\exists m, X_m = j) > 0$ donc $\sum_m \mathbb{P}_i(X_m = j) > 0$ et il existe m tel que $\mathbb{P}_i(X_m = j) > 0$. \square

Corollaire 1.6. — *la relation $i \leftrightarrow j$ est une relation d'équivalence. Les classes d'équivalence sont appelées classes de communication.*

Démonstration. — Pour la transitivité si $i \rightarrow j$ et $j \rightarrow k$ alors il existe m, n tels que $p_{i,j}^{(n)} > 0$ et $p_{j,k}^{(m)} > 0$ donc $p_{i,k}^{(n+m)} \geq p_{i,j}^{(n)} p_{j,k}^{(m)} > 0$. \square

Définition 1.3. — *La chaîne est dite irréductible s'il y a une seule classe. La classe C est dite ouverte s'il existe $i \in C$ et $j \notin C$ tels que $i \rightarrow j$.*

Proposition 1.7. — *La classe C est ouverte ssi il existe $i \in C$ et $j \notin C$ tels que $p_{i,j} > 0$.*

On suppose la classe fermée et que p.s. $X_0 \in C$. Alors presque sûrement $\forall n, X_n \in C$. En d'autres termes une chaîne reste toujours p.s. dans une classe fermée.

Démonstration. — Pour la première partie il existe un chemin fini de C à l'extérieur de C . Il y a bien un moment où il quitte C .

Pour la seconde partie on fait une récurrence et on voit que l'hérédité découle de la remarque suivante. cOMME C est fermée, si $i \in C$, alors pour tout $j \notin C$, $p_{i,j}^{(n)} = 0$ et donc en sommant sur tous les $j \notin C$, $\mathbb{P}_i(X_n \in C^c) = 0$ et donc

$$\mathbb{P}(X_n \notin C) = \sum_i \mathbb{P}(X_0 = i) \mathbb{P}_i(X_n \notin C) = 0$$

On utilise bien sûr le fait que l'intersection d'ensemble de probabilité 1 est de probabilité 1. \square

3.2 Récurrence et transience

Définition 1.4. — *On dit que i est transient si $\sum_n p_{ii}^{(n)} < +\infty$ et récurrent si $\sum_n p_{ii}^{(n)} = +\infty$.*

Proposition 1.8. — *La récurrence est une propriété de classe. Soit C une classe de communication. Alors soit tous les éléments de C sont récurrents, soit ils sont tous transients.*

Démonstration. — soit $i, j \in C$ on a $i \leftrightarrow j$ donc il existe k, m tels que $p_{ij}^{(m)} > 0$ et $p_{ji}^{(k)} > 0$ On déduit de l'inégalité

$$p_{ii}^{(n+m+k)} \geq p_{ij}^{(m)} p_{jj}^{(n)} p_{ji}^{(k)}$$

et de l'inégalité symétrique obtenue en permutant i et j , que les séries $\sum_n p_{ii}^{(n)}$ et $\sum_n p_{jj}^{(n)}$ sont de même nature. \square

Le nombre de retours en i (ou temps passé en i à partir de 1) est $N_i = \sum_{n \geq 1} \mathbf{1}_{(X_n=i)}$ et la probabilité de retour en i est $f_i := \mathbb{P}_i(N_i > 0) = \mathbb{P}_i(T < +\infty)$ avec $T = \inf \{n \geq 1 : X_n = i\}$.

Lemme 1.9. — pour tout $r \in \mathbb{N}$, $\mathbb{P}_i(N_i > r) = f_i^{r+1}$.

Démonstration. — Preuve par récurrence. La propriété est vraie pour $r = 0$. Et si elle est vraie pour r , alors on conditionne par la valeur m de T et on applique la propriété de Markov à l'instant m . \square

Théorème 1.10 (dichotomie). — Si $f_i = 1$ alors presque sûrement $X_n = i$ pour une infinité d'entiers n et i est un état récurrent.

Si $f_i < 1$, alors presque sûrement $X_n = i$ pour seulement un nombre fini d'entiers et l'état i est transient.

Démonstration. — On observe que d'une part

$$\mathbb{E}_i [N_i] = \sum_{k \in \mathbb{N}} \mathbb{P}_i(N_i > k) = \sum_k f_i^{k+1} = \begin{cases} \frac{f_i}{1-f_i} & \text{si } f_i < 1; \\ +\infty & \text{si } f_i = 1; \end{cases}$$

et d'autre part

$$\mathbb{E}_i [N_i] = \sum_n \mathbb{E}_i [\mathbf{1}_{(X_n=i)}] = \sum_n p_{ii}^{(n)}$$

d'où la dichotomie. Si $f_i = 1$ alors $N_i = +\infty$ ps (car pour tout k , $\mathbb{P}_i(N_i > k) = 1$) et i est récurrent, alors que si $f_i < 1$ on a $N_i < +\infty$ ps et i est transient. \square

Remarque. — On montre de la même façon que si C est une classe récurrente et $i, j \in C$ alors \mathbb{P}_i ps, on a $N_j = +\infty$ et $\mathbb{E}_i [N_j] = \sum_n p_{ij}^{(n)} = +\infty$ Si C est une classe transiente et $i, j \in C$ alors \mathbb{P}_i ps, on a $N_j < +\infty$ et $\mathbb{E}_i [N_j] = \sum_n p_{ij}^{(n)} < +\infty$

Proposition 1.11. — Soit C une classe de communication.

1. Si C est ouverte alors C est transiente. et p.s. il existe $n_0 = n_0(\omega)$ tel que $\forall n \geq n_0, X_n \notin C$ (au bout d'un moment on sort définitivement d'une classe ouverte finie).
2. Si C est fermée alors C est récurrente.

Démonstration. — 1) Soit $i \in C, j \notin C$ tels que $p_{ij} > 0$ On a $j \not\rightarrow i$ donc, $1 = \mathbb{P}_j(\forall n \geq 0, X_n \neq i)$. Donc, par Markov,

$$\mathbb{P}_i(N_i = 0) \geq \mathbb{P}_i(X_1 = j, \forall n \geq 1 X_n \neq i) = p_{ij} \mathbb{P}_j(\forall n \geq 0, X_n \neq i) = p_{ij} > 0$$

i.e. $f_i = \mathbb{P}_i(N_i > 0) < 1$ et i est transient.

On note $N_C = \sum_{j \in C} N_j$ le temps passé en C , et on a $\mathbb{E}_i[N_C] = \sum_{j \in C} \mathbb{E}_i[N_j]$ qui est une somme finie. Or pour tout $j \in C$ on a $\mathbb{E}_i[N_j] < +\infty$ donc $\mathbb{E}_i[N_C] < +\infty$ et donc $N_C < +\infty$ ps.

2) C est finie et fermée. Soit $i \in C$, sous $\mathbb{P}_i, X_0 = i$ donc on a vu que ps $\forall n, X_n \in C$ et donc $N_C = +\infty$ ps, donc $\mathbb{E}_i[N_C] = +\infty$ et donc il existe j tel que $\mathbb{E}_i[N_j] = +\infty$, donc i est récurrent. \square

Corollaire 1.12. — Une chaîne de Markov irréductible (sur un espace d'états fini) est récurrente.

Remarque. — Sur un espace d'états infini tout peut arriver. Sur \mathbb{Z} la marche aléatoire simple est récurrente (TCL) alors que la marche aléatoire biaisée est transiente car par la loi forte des grands nombres, p.s. $X_n \rightarrow +\infty$

Corollaire 1.13. — Une chaîne de Markov sur un espaces d'états fini se retrouve au bout d'un moment (aléatoire) dans une classe fermée dont elle ne sort pas, et dont elle visite tous les points une infinité de fois.

SECTION 4

Temps d'atteinte et probabilités d'absorption

Exemple 1.6. — Sur l'espace d'états $\{1, 2, 3, 4\}$, $p_{i,i+1} = 0.4 = 1 - p_{i,i-1}$ si $i = 2, 3, 1$ et 4 absorbants.

Si 1 est gagnant et 4 perdant et que l'on me donne deux fois ma mise quand je pars de 3, et une seule fois quand je pars de 2, quelle doit être ma stratégie ? (on part de 3 car $20.473 > 0.789$).

On considère une chaîne de Markov, à espace d'états fini, non irréductible. On sait que la chaîne finira par se fixer dans une classe fermée. Soit A une réunion de classes fermées. On veut déterminer la probabilité que la chaîne se fixe dans A .

Proposition 1.14. — Soit $\sigma_A = \inf \{n \geq 0 : X_n \in A\}$ avec la convention $\inf \emptyset = +\infty$.

La famille des probabilités d'absorption $h(i) = \mathbb{P}_i(\sigma_A < +\infty)$ satisfait

$$h(i) = \begin{cases} 1 & \text{si } i \in A ; \\ \sum_j p_{ij} h(j) & \text{sinon.} \end{cases}$$

Démonstration. — On conditionne par X_1 si $i \notin A$ et on utilise la propriété de Markov. Plus précisément, si $(\tilde{X}_n = X_{1+n}, n \geq 0)$, on a sous \mathbb{P}_i , comme $i \notin A$,

$$\sigma_A = 1 + \sigma_A(\tilde{X}).$$

En conséquence,

$$\begin{aligned} h(i) &= \sum_j \mathbb{P}(\sigma_A < +\infty \mid X_1 = j) \mathbb{P}_i(X_1 = j) \\ &= \sum_j p_{ij} \mathbb{P}(\sigma_A(\tilde{X}) < +\infty \mid X_1 = j) \\ &= \sum_j p_{ij} \mathbb{P}_j(\sigma_A < +\infty) = \sum_j p_{ij} h(j). \end{aligned}$$

□

Observons que dans la preuve précédente A est un ensemble quelconque. On munit \mathbb{R}^I de la relation d'ordre partiel $h \leq k$ si $\forall i \in I, h(i) \leq k(i)$. On dit que h est un élément minimal de $A \subset \mathbb{R}^I$ si c'est un minorant de A . On a le résultat suivant, qui entraîne l'unicité, et dont la preuve est admise.

Proposition 1.15. — *La famille des probabilités d'absorption est minimale parmi la famille des vecteurs λ de nombres positifs qui vérifient*

$$\lambda(i) = \begin{cases} 1 & \text{si } i \in A ; \\ \sum_j p_{ij} \lambda(j) & \text{sinon.} \end{cases}$$

Lorsque A est la réunion de toutes les classes fermées, on peut se poser la question de l'évaluation du temps moyen d'absorption dans A : $k(i) := \mathbb{E}_i[\sigma_A]$. En conditionnant par X_1 on montre que

Proposition 1.16. — *La famille des temps moyen d'absorption vérifie*

$$h(i) = \begin{cases} 0 & \text{si } i \in A ; \\ 1 + \sum_j p_{ij} h(j) & \text{sinon.} \end{cases}$$

Observons que le résultat est vrai quel que soit l'ensemble A . On peut également montrer une propriété de minimalité que nous n'énoncerons pas ici. On a parfois besoin de connaître la loi du temps de sortie d'un point.

Lemme 1.17. — *On suppose que la chaîne est issue de i et que i est non absorbant. Soit $T = \inf n \geq 1 : X_n \neq i$ Alors sous \mathbb{P}_i , les variables aléatoires X_T et T sont indépendantes, T de loi géométrique de paramètre $1 - p_{ii}$ et X_T de loi*

$$\mathbb{P}(X_T = j) = \frac{p_{ij}}{1 - p_{ii}}$$

Exemple 1.7. — On revient sur l'exemple 3; Partant de 0 la proba de toucher 6 est donc la proba $\mathbb{P}_0(X_T = 4) = 1/4$ et la probabilité d'aller dans la classe fermée $\{1, 2, 3\}$ est $3/4$. Le temps pour toucher 3 peut être $+\infty$ si je suis absorbé dans la classe fermée $\{4, 5, 6\}$. Donc la moyenne de ce temps est $+\infty$.

SECTION 5

Mesures Invariantes

On dit que la mesure positive λ est invariante (ou encore stationnaire) si $\lambda P = \lambda$.

Proposition 1.18. — Si $(X_n)_{n \in \mathbb{N}}$ est Markov (λ, P) et λ est invariante, alors la loi de X_n est constante et vaut λ . En outre pour tout m , $(X_{m+n}, n \geq 0)$ est Markov (λ, P)

Démonstration. — En effet $X_n \sim \lambda P^n = \lambda$. En outre, par la propriété de Markov,

$$\begin{aligned} \mathbb{P}((X_{m+n}, n \geq 0) \in A) &= \sum_i \mathbb{P}((X_{m+n}, n \geq 0) \in A \mid X_m = i) \mathbb{P}(X_m = i) \\ &= \sum_i \mathbb{P}_i((X_n, n \geq 0) \in A) \mathbb{P}(X_m = i) \\ &= \sum_i \mathbb{P}_i((X_n, n \geq 0) \in A) \lambda(i) \\ &= \mathbb{P}((X_n, n \geq 0) \in A) \end{aligned}$$

car si cette quantité ne dépend pas de m elle égale sa valeur quand $m = 0$. \square

Les limites des lois de X_n sont automatiquement des probabilités invariantes sur un espace fini.

Proposition 1.19. — On suppose I fini et que pour un $i_0 \in I$ on ait

$$\forall j, \quad p_{i_0, j}^{(n)} \rightarrow \pi_j$$

Alors π est une probabilité invariante.

Démonstration. — Par sommation des limites

$$\pi_j = \lim p_{i_0, j}^{(n)} = \lim p_{i_0, j}^{(n+1)} = \lim \sum_k p_{i_0, k}^{(n)} p_{kj} = \sum_k \pi_k p_{kj} = (\pi P)_j$$

En outre, $\pi_j \geq 0$ car limite d'une suite de nombres positifs et

$$1 = \sum_j p_{i_0, j}^{(n)} \rightarrow \sum_j \pi_j$$

donc π est bien une probabilité. \square

Proposition 1.20. — Soit X une chaîne de Markov irréductible sur un espace d'états fini. Alors la chaîne admet une unique probabilité invariante π . Pour tout i , $\pi(i) > 0$ et en outre

$$\pi(i) = \frac{1}{\mathbb{E}_i[T_i]}, \quad \text{avec } T_i = \inf \{n \geq 1 : X_n = i\}.$$

Démonstration. — Admise □

5.1 Théorème ergodique

Théorème 1.21. — Soit X une chaîne irréductible. Alors, pour toute fonction f , presque sûrement

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \pi(f)$$

Démonstration. — (Admise) □

Exemple 1.8. — Pour l'exemple de l'introduction, la proportion de temps passé en temps long dans l'état 2, en partant de 1, est de $3/8$. Elle vaut $\pi(2) = 3/8$ avec π unique proba invariante de la chaîne réduite à $\{1, 2, 3\}$.

5.2 Mesures réversibles Il est parfois difficile de calculer la mesure de proba invariante même quand on sait qu'elle existe et est unique.

Définition 1.5. — On dit que la mesure π est réversible si elle vérifie les équations de bilan détaillé

$$\pi(x)p_{xy} = \pi(y)p_{yx}$$

Lemme 1.22. — Une mesure réversible est invariante.

Démonstration. —

$$\pi P(j) = \sum_i \pi(i)p_{ij} = \sum_i \pi(j)p_{ji} = \pi(j) \sum_i p_{ji} = \pi(j)$$

□

Remarque. — Si la matrice de transition est symétrique, alors la mesure $\pi(i) = 1$ est réversible donc invariante et si on est sur un espace fini, cela entraîne que l'unique mesure invariante est la loi uniforme.

Remarque. — On montre facilement que si $0 \leq N$, et $Y_N = X_{N-n}, 0 \leq n \leq N$ est la chaîne retournée dans le temps (à l'instant N), et si $X_0 \sim \pi$ loi réversible, alors Y est une chaîne de Markov (π, P) en montrant que

$$\mathbb{P}(Y_0 = i_0, \dots, Y_N = i_N) = \mathbb{P}(X_0 = i_0, \dots, X_N = i_N)$$

5.3 Chaînes non irréductibles On admet le résultat suivant de théorie de l'intégration:

Proposition 1.23. — Soit X_n une suite de variables aléatoires discrètes convergeant presque sûrement vers une variable aléatoire X .

1. si X_n est à valeurs dans l'espace fini I , alors pour tout $i \in I$ on a

$$\mathbb{P}(X_n = i) \rightarrow \mathbb{P}(X = i)$$

et pour toute fonction $f : I \rightarrow \mathbb{R}$,

$$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)].$$

2. Si X_n, X sont réelles, alors pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée:

$$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)].$$

Définition 1.6. — On dit que la mesure λ ne charge pas l'ensemble A si $\lambda(A) = 0$.

Proposition 1.24 (Les mesures invariantes ne chargent pas les classes ouvertes)

Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur un espace d'états fini I . Soit π une mesure invariante. Alors si C est une classe ouverte $\pi(C) = 0$.

Démonstration. — On pose $Z_n(\omega) = \mathbf{1}_{(X_n(\omega) \in C)}$. Alors, par l'étude du comportement asymptotique des chaînes de Markov, presque sûrement il existe $n_0 = n_0(\omega)$ tel que pour tout $n \geq n_0$, $X_n \notin C$. Cela entraîne que presque sûrement $Z_n(\omega) \rightarrow Z(\omega) = 0$. Donc, par la proposition précédente,

$$\mathbb{P}(X_n \in C) = \mathbb{E}[Z_n] \rightarrow \mathbb{E}[Z] = 0.$$

Il suffit maintenant d'imposer comme loi initiale $X_0 \sim \pi$. Alors on sait que pour tout n , $X_n \sim \pi$ et donc $\mathbb{P}(X_n \in C) = \pi(C)$. \square

On a une description assez précise de l'ensemble des mesures invariantes. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur un espace d'états fini I . Si C est une classe fermée, alors

$$\forall i \in C, \forall j \notin C, p_{ij} = 0$$

et donc

$$\forall i \in C, \sum_{j \in C} p_{ij} = 1$$

On peut donc considérer la chaîne de Markov X^C qui est restreinte à C : elle a pour matrice de transition P^C définie par

$$P_{ij}^C = p_{ij} \quad (i, j \in C)$$

Elle est irréductible et admet donc une unique mesure invariante π^C . En complétant avec des zéros, on considère également π^C comme une mesure de probabilité sur I tout entier: $\pi^C(i) = 0$ si $i \notin C$.

Proposition 1.25. — π est une mesure invariante ssi il existe une probabilité $(\alpha_C)_C$ fermée telle que

$$\pi = \sum_{C \text{ fermée}} \alpha_C \pi^C.$$

Démonstration. — Vérifions qu'une telle mesure est invariante. Par linéarité il suffit de vérifier que $\pi^C P = \pi^C$ et ceci découle de la définition de π^C : si $j \notin C$,

$$\pi^C(j) = 0, (\pi^C P)(j) = \sum_i \pi^C(i) p_{ij} = 0$$

car si $i \notin C$, $\pi^C(i) = 0$ et si $i \in C$, $p_{ij} = 0$ (car $j \notin C$). Si $j \in C$,

$$(\pi^C P)(j) = \sum_i \pi^C(i) p_{ij} = \sum_{i \in C} \pi^C(i) p_{ij} = (\pi^C P^C)(j) = \pi^C(j).$$

Réciproquement, soit π une mesure invariante. Soit C une classe fermée telle que $\pi(C) > 0$. On pose $(\mu(i) = \frac{1}{\pi(C)} \pi(i))_{i \in C}$ C'est une probabilité sur C . Nous allons montrer que $\mu P^C = \mu$. Par unicité de la mesure invariante pour la chaîne irréductible restreinte à C , cela entraînera que $\mu = \pi^C$ et le résultat de la proposition.

Soit $j \in C$, Comme on a $\mu(j) \geq (\mu P^C)(j)$ car

$$\begin{aligned} (\mu P^C)(j) &= \sum_{i \in C} \mu(i) P_{ij}^C = \frac{1}{\pi(C)} \sum_{i \in C} \pi(i) p_{ij} \\ &\leq \frac{1}{\pi(C)} \sum_{i \in I} \pi(i) p_{ij} = \frac{1}{\pi(C)} (\pi P)(j) = \frac{1}{\pi(C)} \pi(j) = \mu(j). \end{aligned}$$

Or μ et μP^C sont des probabilités sur C , donc

$$1 = \sum_{j \in C} \mu(j) \geq \sum_{j \in C} \mu P^C(j) = 1$$

Il y a donc égalité pour tout $j \in C$. □

Convergence vers l'équilibre

La matrice de transition $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ vérifie $P^{2n} = I$ et $P^{2n+1} = P$. En conséquence $p_{ij}^{(n)}$ ne converge pas

Définition 1.7. — La période de l'état i est

$$d(i) = \text{pgcd}\{n \geq 1, p_{ii}^{(n)} > 0\}.$$

On dit que l'état i est apériodique si $d(i) = 1$.

Proposition 1.26. — La période est une propriété de classe de communication.

Démonstration. — On note $A_i := \{n \geq 1, p_{ii}^{(n)} > 0\}$. Si i et j communiquent alors il existe $k, l \geq 0$ tels que $p_{ij}^{(k)} p_{ji}^{(l)} > 0$. Alors

$$p_{ii}^{(k+l)} \geq p_{ij}^{(k)} p_{ji}^{(l)} > 0$$

donc $k+l \in A_i$ et $d(i) \mid k+l$. Alors pour tout $n \in A_j$

$$p_{ii}^{(n+k+l)} \geq p_{ij}^{(k)} p_{jj}^{(n)} p_{ji}^{(l)} > 0$$

donc $k+l+n \in A_i$ donc $d(i) \mid k+l+n$ et comme $d(i) \mid k+l$ on a $d(i) \mid n$. Donc $d(i) \mid d(j)$. De même $d(j) \mid d(i)$ et donc $d(i) = d(j)$. \square

Proposition 1.27. — On considère une chaîne irréductible. Si i est apériodique, alors tous les états sont apériodiques. En outre, comme I est fini, il existe n_0 tel que pour tout $n \geq n_0$, et tous j, k $p^{(n)}(j, k) > 0$.

Démonstration. — On pose $S = \{n \geq 1, p_{ii}^{(n)} > 0\}$ Alors S est un semi groupe de $(\mathbb{N}, +)$ i.e. $S + S \subset S$ et $\text{pgcd}(S) = 1$, donc si $I = S - S$, alors I est un sous groupe de $(\mathbb{Z}, +)$ donc de la forme $d\mathbb{Z}$ et comme $\text{pgcd}(S) = 1$ on a $d = 1$ donc $1 \in S - S$: il existe $n_0 \in S$ tel que $n_0 + 1 \in S$ Alors si $k \geq n_0^2 + n_0$, quand on fait la division de k par n_0 , $k = qn_0 + r$, on a $q \geq n_0$ donc $k = r(n_0 + 1) + (q - r)n_0 \in S$. Comme $i \rightarrow j$ on en déduit qu'il existe $n(i, j)$ tel que pour tout $n \geq n(i, j)$ $p_{i,j}^{(n)} > 0$. \square

Théorème 1.28. — Soit une chaîne de Markov irréductible de proba invariante π et apériodique. Alors, pour toute loi initiale

$$\mathbb{P}(X_n = i) \rightarrow \pi(i)$$

en particulier pour tout i ,

$$p_{ij}^{(n)} \rightarrow \pi(j),$$

ce que l'on peut écrire

$$P^n \rightarrow P^\infty$$

avec P^∞ la matrice dont toutes les lignes sont égales à π .

Démonstration. — Preuve admise. □

Remarque. — Sous les hypothèses précédentes on peut même prouver beaucoup mieux (c'est une partie du théorème de Perron Fröbenius)

1. toutes les autres valeurs propres de P sont de module strictement inférieur à 1 :

$$\rho = \sup \{|\lambda|, \lambda \in \text{sp}(P), \lambda \neq 1\} \in [0, 1[.$$

2. La matrice P^n converge exponentiellement vite vers P^∞ : il existe $C > 0$ telle que

$$\|P^n - P^\infty\| \leq C\rho^n.$$

3. Si x est un vecteur non nul à coordonnées positives et qu'il existe λ avec $Px = \lambda x$, alors $\lambda = 1$ et x est constant.

Introduction à la Statistique

On observe des réalisations $(x_1 = X_1(\omega), \dots, x_n = X_n(\omega))$ de variables aléatoires. Par exemple:

1. (X_1, \dots, X_n) sont IID, Independent Identically Distributed, c'est à dire indépendantes de même loi P_X . On parle parfois d'un échantillon de taille n .
2. $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de matrice de transition P .

Le but de *l'estimation* est d'apprendre certaines caractéristiques, de la loi P_X dans le cas d'un échantillon et de la matrice P dans le cas d'une chaîne de Markov, à partir du vecteur des observations (x_1, \dots, x_n) .

Le but de la *prévision* est de prévoir x_{n+1} à partir des observations. Une *statistique* est une fonction des observations.

Les *grandeurs théoriques* sont celles qui dépendent de la loi P_X (ou de la matrice P). Par exemple : $\mathbb{E}[X_1]$.

Les *grandeurs empiriques* sont celles qui dépendent des observations. Par exemple la moyenne empirique:

$$\bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i.$$

Dans le cas d'un échantillon, la loi forte des grands nombres nous dit que \bar{x}_n est proche de $\mathbb{E}[X_1]$ si n est grand.

Statistique descriptive

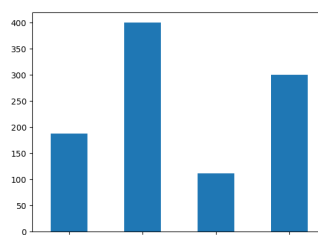
1.1 Variables qualitatives Une variable qualitative, ou *facteur*, est observée comme une catégorie, un code par exemple : sexe, nationalité, mention au Bac. Les différentes valeurs possibles sont appelées *modalités*

L'observation de la variable A , de modalités A_1, \dots, A_k peut se résumer par un *tableau des fréquences*

Modalités de A	A_1	A_2	\dots	A_n
Effectifs observés	n_1	n_2	\dots	n_k
Fréquences observées	f_1	f_2	\dots	f_k

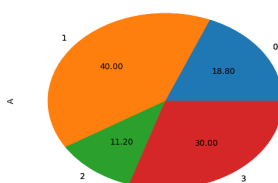
On a évidemment $\sum_i n_i = n$ et $\sum_i f_i = 1$ car $f_i = \frac{n_i}{n}$.
Les représentations graphiques les plus courantes sont:

1. Le diagramme en barres : la hauteur de chaque barre est proportionnelle



à la fréquence (l'effectif) de chaque modalité.

2. Le diagramme circulaire (dit en camembert): l'aire du secteur est pro-



portionnelle à la fréquence.

Si on observe un n -échantillon, la loi forte des grands nombres dit que lorsque n est grand, les fréquences observées sont proches des probabilités d'obtenir la modalité: presque sûrement

$$f_i = \frac{n_i}{n} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{(X_k=i)} \rightarrow \mathbb{E} [\mathbf{1}_{(X_1=i)}] = \mathbb{P}(X_1 = i)$$

La prédiction de x_{n+1} est obtenue en tirant une variable aléatoire qui a comme probabilité d'obtenir une modalité la fréquence d'apparition.

1.2 Variables Quantitatives L'observation d'une variable quantitative est une quantité mesurée, par exemple âge, salaire, note, taux de sucre ... Nous ne considérerons que des variables quantitatives discrètes. On note (x_1, \dots, x_n) l'échantillon des valeurs numériques

1.2.1 *Résumés numériques* Les mesures de position

1. Le *mode* est la valeur la plus fréquente de l'échantillon
2. La *moyenne empirique* est $\bar{x}_n = \frac{1}{n} \sum_{k=1}^n x_k$
3. Une *médiane* est un nombre m séparant l'échantillon ordonné

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

en 2 parties:

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(x_i \leq m)} \geq \frac{1}{2}, \quad \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(x_i \geq m)} \geq \frac{1}{2}$$

Si $n = 2k + 1$, la médiane est unique $m = x_{(k+1)}$.

Si $n = 2k$, tout nombre de $[x_{(k)}, x_{(k+1)}]$ et on choisit arbitrairement

$$m = \frac{1}{2}(x_{(k)} + x_{(k+1)})$$

La médiane est plus robuste aux valeurs extrêmes (outliers) que la moyenne.

4. Le *quantile* $Q(p)$ d'ordre $p \in [0, 1]$ sépare à peu près l'ensemble des observations en deux parties de tailles respectives np et $n(1 - p)$. Plus précisément on doit avoir

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(x_i \leq Q(p))} \geq p, \quad \text{et} \quad \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(x_i \geq Q(p))} \geq 1 - p$$

Si $\lfloor \cdot \rfloor$ et $\lceil \cdot \rceil$ désigne les parties entières inférieures et supérieures ($\lceil u \rceil - 1 < u \leq \lceil u \rceil$), alors $Q(p)$ est un nombre de $[x_{(\lceil np \rceil)}, x_{(\lfloor np \rfloor + 1)}]$. On choisit usuellement $Q(p) = x_{(\lceil np \rceil)}$ si np n'est pas un entier et $Q(p) = \frac{1}{2}(x_{(\lceil np \rceil)} + x_{(\lfloor np \rfloor + 1)})$ sinon. Certains utilisent cette formule même quand np est un entier. Il y a d'autres choix d'interpolation des quantiles qui dépendent du logiciel utilisé : par exemple python avec pandas et excel utilisent des interpolations différentes. Il est donc important de préciser la formule utilisée.

Pour $p = 0.25, 0.5, 0.75$ on parle de *quartiles* Pour $p = 0.1, 0.2, \dots, 0.9$ on parle de *déciles*

Les mesures de dispersion

- l'étendue est le maximum moins le minimum $x_{(n)} - x_{(1)}$.
- la variance empirique est $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum x_i^2 - \bar{x}_n^2$.
- L'écart type est σ_n .
- la variance corrigée est $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$. L'écart type corrigé (standard deviation en anglais) est S_n .
- L'écart interquartile est $Q(0.75) - Q(0.25)$.
- L'écart moyen absolu est $EMA = \frac{1}{n} \sum_i |X_i - m|$

1.2.2 Résumés graphiques

La boîte à moustaches ou boxplot : . Le rectangle central est délimité par le premier et le troisième quartile et la médiane y est symbolisée par un trait. Les "moustaches" partent de chaque côté jusqu'à la valeur minimale et maximale de l'échantillon, sous réserve que leur longueur ne dépasse pas $1.5(Q(0.75) - Q(0.25))$, soit 1.5 fois la hauteur de la boîte. Sinon, les moustaches s'arrêtent à la dernière valeur avant cette limite et les valeurs restantes sont représentées par des points isolés.

```
df = pd.DataFrame(np.random.randint(low=0,high=25, size=(10, 5)),
                  columns=['A', 'B', 'C', 'D', 'E'])
```

```
df.plot(kind='box')
```

```
df
   A  B  C  D  E
0 17 17 16 10  1
1  4  3 24  9 16
2 22 23 13 18 14
3 18 22  5 13  9
4 22  9 14  2 14
5 15 10  6  9  6
6 18 16  5 19 23
7  2 16 19  0  8
8 13  5 24  9 14
9 19  3 20  6 13
```

```
df.median()
```

```
A    17.5
B    13.0
C    15.0
D     9.0
E    13.5
```

```
df.mean()
```



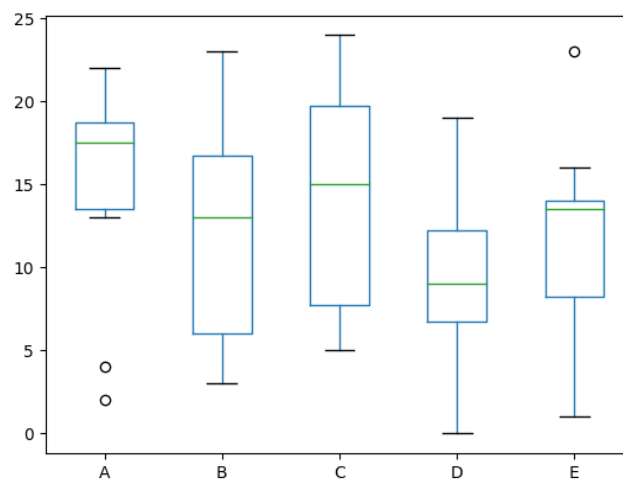
```
A    15.0
B    12.4
C    14.6
D     9.5
E    11.8
```

```
df.quantile([0.25,0.5,0.75],interpolation='midpoint')
```

```
      A      B      C      D      E
0.25 14.0   7.0   9.5   7.5   8.5
0.50 17.5  13.0  15.0   9.0  13.5
0.75 18.5  16.5  19.5  11.5  14.0
```

```
df.quantile([0.25,0.5,0.75])
```

```
      A      B      C      D      E
0.25 13.50   6.00   7.75   6.75   8.25
0.50 17.50  13.00  15.00   9.00  13.50
0.75 18.75  16.75  19.75  12.25  14.00
```



SECTION 2

Estimation paramétrique

2.1 Estimation de la probabilité de succès dans une loi binomiale

On suppose que l'on interroge $n = 500$ personnes sur leur intention de vote et

que 375 disent qu'elles vont voter pour le candidat A et 125 pour le candidat B . On estime la probabilité de voter pour A par la fréquence $375/500 = 75\%$. On suppose que (X_1, \dots, X_n) sont IID de loi $\mathcal{B}(1, p)$ i.e.

$$\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0) = p \in]0, 1[.$$

Le paramètre p est inconnu et on veut l'estimer.

Alors $S_n = X_1 + \dots + X_n \sim \mathcal{B}(np)$ et $\mathbb{E}[S_n] = np$, $\text{Var}(S_n) = np(1-p)$.

par la loi forte des grands nombres, presque sûrement la moyenne empirique converge vers l'espérance

$$\bar{X}_n = \frac{1}{n}S_n \rightarrow \mathbb{E}[X_1] = p \quad p.s.$$

On définit $\hat{p}_n = \bar{x}_n = X_n(\omega)$ comme estimateur de la probabilité inconnue p . On a $\mathbb{E}[\hat{p}_n] = p$: il est sans biais.

2.2 Critères de qualité d'un estimateur La famille $(P_\theta, \theta \in \Theta)$ est une famille paramétrée de probabilités sur un ensemble discret S . Par exemple dans le schéma binomial $P_\theta = \mathcal{B}(p)$ avec $p = \theta \in (0, 1)$ le paramètre.

On suppose que l'on dispose d'un n -échantillon (X_1, \dots, X_n) de loi P_θ , c'est à dire d'une suite IID de loi P_θ

Définition 2.1. — Un estimateur $\hat{\theta}_n$ est une variable aléatoire qui est fonction des observations :

$$\hat{\theta}_n(\omega) = g(x_1, \dots, x_n) = g(X_1(\omega), \dots, X_n(\omega)).$$

Définition 2.2. — On dit que l'estimateur $\hat{\theta}_n$ du paramètre θ est sans biais si pour tout $\theta \in \Theta$, $\hat{\theta}_n$ est intégrable et

$$\mathbb{E}_\theta[\hat{\theta}_n] = \theta \quad (\forall \theta \in \Theta).$$

La variance étant une mesure de la dispersion, voire par exemple l'égalité de Tchebyshev, on peut comparer des estimateurs.

Définition 2.3. — L'estimateur $\hat{\theta}_n$ du paramètre θ est quadratiquement uniformément meilleur que l'estimateur $\tilde{\theta}_n$ si

$$\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \leq \mathbb{E}_\theta[(\tilde{\theta}_n - \theta)^2] \quad (\forall \theta \in \Theta).$$

Pour des estimateurs sans biais cela revient à comparer les variances.

On suppose que l'on peut considérer des tailles n de plus en plus grandes d'échantillon.

Définition 2.4. — On dit que l'estimateur $\hat{\theta}_n$ du paramètre θ est (fortement) consistant si pour tout $\theta \in \Theta$

$$\hat{\theta}_n \rightarrow \theta \quad ps$$

Définition 2.5. — On dit que l'estimateur $\hat{\theta}_n$ du paramètre θ converge en moyenne quadratique si

$$\mathbb{E}_\theta \left[(\hat{\theta}_n - \theta)^2 \right] \rightarrow 0.$$

Pour le schéma binomial, \hat{p}_n est un estimateur fortement consistant, sans biais, qui converge en moyenne quadratique. On peut même montrer qu'il est quadratiquement uniformément meilleur parmi tous les estimateurs sans biais.

2.3 Estimateur des moments

Proposition 2.1. — La moyenne empirique

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$$

est un estimateur sans biais, consistant de la moyenne $\mathbb{E}_\theta [X_1]$

Démonstration. — C'est la loi forte des grands nombres, où l'on a supposé X_1 intégrable sous P_θ . \square

Proposition 2.2. — Si $n \geq 2$, et que X_1 est de carré intégrable, alors la moyenne empirique est un estimateur convergent en moyenne quadratique de la moyenne $\mathbb{E}_\theta [X_1]$.

Proposition 2.3. — Si $n \geq 2$, et que X_1 est de carré intégrable, la variance empirique modifiée

$$V_n := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

est un estimateur consistant et sans biais de la variance de X_1 , $\text{Var}_\theta(X_1)$.

Démonstration. — La variance empirique est:

$$\hat{\sigma}_n^2 := \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

est par la loi forte des grands nombres appliquée à X_n , puis X_n^2 , un estimateur fortement consistant de

$$\text{Var}(X) = \mathbb{E}_\theta [X_1^2] - \mathbb{E}_\theta [X_1]^2.$$

Un simple calcul montre que $\mathbb{E}_\theta [\hat{\sigma}_n^2] = \frac{n-1}{n} \text{Var}_\theta(X_1)$. \square

Exemple 2.1. — On considère un n échantillon d'une loi géométrique de paramètre $p \in (0, 1)$. Alors $\bar{X}_n \rightarrow \mathbb{E}_p [X_1] = \frac{1}{p}$ presque sûrement, donc $\hat{p}_n = \frac{1}{\bar{X}_n}$ est un estimateur consistant de p . En revanche, il n'est pas sans biais.

Application numérique : on tire 100 géométriques de paramètre $p = 0,3$ et on obtient $\hat{p}_{100} \simeq 0,306$.

2.4 Intervalles de confiance

Définition 2.6. — Un intervalle de confiance de niveau $a \in (0, 1)$ est un intervalle aléatoire I_n qui ne dépend que des observations (X_1, \dots, X_n) , mais pas du paramètre θ , et tel que

$$\mathbb{P}_\theta(\theta \in I_n) \geq a \quad (\forall \theta \in \Theta).$$

On peut comparer deux intervalles de confiance de même niveau, en disant que le plus petit est le meilleur.

Définition 2.7. — La suite d'intervalles de confiance $(I_n)_n$ est de niveau asymptotique a si

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta(\theta \in I_n) = a \quad (\forall \theta \in \Theta).$$

En terminale, on apprend qu'un intervalle de confiance correct lorsque n est grand, pour le paramètre p d'une binomiale est

$$I_n(\text{terminale}) = \left[\hat{p}_n - \frac{1}{\sqrt{n}}, \hat{p}_n + \frac{1}{\sqrt{n}} \right].$$

On note q_α les quantiles d'ordre α de la loi normale centrée réduite $\mathcal{N}(0, 1)$. On a le résultat suivant

Proposition 2.4. — Pour le modèle binomial

$$I_n(1 - \alpha) = \left[\hat{p}_n - q_{1-\alpha/2} \sqrt{\hat{p}_n(1 - \hat{p}_n)} \frac{1}{\sqrt{n}}, \hat{p}_n + q_{1-\alpha/2} \sqrt{\hat{p}_n(1 - \hat{p}_n)} \frac{1}{\sqrt{n}} \right]$$

est une suite d'intervalles de confiance de niveau asymptotique $1 - \alpha$.

On a, pour $\alpha = 0.05$, $q_{0.975} \simeq 1.96 \simeq 2$ et si $p \in [0.2, 0.8]$, $\sqrt{p(1-p)} \sim 0.5$ et donc

$$I_n(\text{terminale}) \simeq I_n(0.95)$$

est un intervalle asymptotique de niveau 95%. On note aussi que $q_{0.95} \sim 1.65$, $q_{0.995} \sim 2.58$.

En pratique on considère que l'on peut utiliser l'intervalle de confiance précédent pour la loi binomiale si:

$$n \geq 30, np \geq 5, n(1-p) \geq 5$$

Il faut donc explicitement vérifier ces conditions lors de l'application de la proposition.

Proposition 2.5. — Dans le cas d'un échantillon d'une variable de carré intégrable, on a un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour $\mathbb{E}_\theta [X_1]$ en prenant

$$I_n(1 - \alpha) = \left[\bar{X}_n - q_{1-\alpha/2} n^{-1/2} s_n, \bar{X}_n + n^{-1/2} q_{1-\alpha/2} s_n \right]$$

en prenant $s_n = \sqrt{V_n}$ ou $s_n = \hat{\sigma}_n$.

Démonstration. — Le théorème central limite, et ses extensions. □

Pour l'exemple des 100 géométriques, on a observé $\hat{\sigma}_n \simeq 2.2$, ce qui donne un intervalle de confiance, pour $\frac{1}{p}$,

$$I(0.95) \simeq [3.27 \pm 1.96 * 2.2/10] \simeq [2.84, 3.7]$$

et donc pour p :

$$J \simeq [0.27, 0.35]$$

2.5 Estimation de la matrice de transition d'une chaîne de Markov

On se donne une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de matrice de transition P . Pour estimer p_{ij} on compte, sur une longue période, le nombre de transitions de i à j divisé par le nombre de passages en i .

Proposition 2.6. — Si la chaîne est irréductible, alors

$$\hat{\theta}_n = \frac{\sum_{k=0}^n \mathbf{1}_{(X_k=i, X_{k+1}=j)}}{\sum_{k=0}^n \mathbf{1}_{(X_k=i)}}$$

est une suite d'estimateurs consistant de p_{ij} .

Démonstration. — Une démonstration hors de portée de la $L2$, mathématiquement, mais facile à comprendre conceptuellement, est de considérer que les excursions hors de i forme une suite *IID* sous \mathbb{P}_i , et donc si on note $(T_k)_{k \in \mathbb{N}}$ les temps de passage successifs en i , alors les variables aléatoires $\xi_k = \mathbf{1}_{(X_{1+T_k}=j)}$ sont *IID* de loi $\mathbb{P}(\xi_k = 1) = p_{ij}$ et donc on se retrouve dans un modèle binomial avec a peu près $N_n = \sum_{k \geq 1} \mathbf{1}_{(T_k \leq n)} \sim n\pi_i$ variables indépendantes

Une autre preuve, moins conceptuelle, consiste à montrer que $(Y_n = (X_n, X_{n+1}), n \in \mathbb{N})$ est une chaîne de Markov irréductible sur I^2 de mesure invariante, si π est la proba invariante de P ,

$$\lambda(i, j) = \pi(i)p_{ij}$$

En conséquence le théorème ergodique donne la convergence presque sûre

$$\frac{1}{n} \sum_{k=0}^n \mathbf{1}_{(Y_k=(i,j))} \rightarrow \lambda(i, j)$$

que l'on combine avec

$$\frac{1}{n} \sum_{k=0}^n \mathbf{1}_{(X_k=i)} \rightarrow \pi(i)$$

□

SECTION 3

Tests

3.1 Introduction

Exemple 2.2. — On s'intéresse à la répartition du sexe des enfants nouveaux-nés en garçons. On dispose des résultats d'un sondage, selon lequel sur 429440 naissances, 60n a dénombré 221023 filles. On se demande si cette répartition entre filles et garçons est compatible avec l'hypothèse d'équiprobabilité de naissance des garçons et des filles.

On est donc dans un schéma binomial de paramètre $p \in (0, 1)$ et on désire tester l'hypothèse " $p = 0.5$ " contre l'hypothèse " $p \neq 0.5$ ".

On suppose que l'espace des paramètres Θ est la réunion de deux parties disjointes Θ_0 et Θ_1 . On désire, à partir d'observations (s_1, \dots, x_n) décider si la valeur inconnue du paramètre θ est dans Θ_0 ou dans Θ_1 .

- l'hypothèse $\{\theta \in \Theta_0\}$, notée encore H_0 , est l'hypothèse nulle.
- l'hypothèse $\{\theta \in \Theta_1\}$, notée H_1 , est l'hypothèse alternative ou contre hypothèse.

Si $H_0 = \{\theta_0\}$ (resp. $H_1 = \{\theta_1\}$) on dit que l'hypothèse nulle est simple (resp. l'hypothèse alternative est simple).

On dit également que l'on teste H_0 contre H_1 .

On peut commettre deux types d'erreur:

Type I : rejeter H_0 alors que $\theta \in \Theta_0$: c'est l'erreur de première espèce.

Type II : accepter H_0 alors qu'en fait $\theta \in \Theta_1$: c'est l'erreur de seconde espèce.

3.2 Zone de rejet Etant données les observations (x_1, \dots, x_n) avec $x_i \in E$, on fabrique une zone de rejet, non aléatoire,

$$R \subset E^n$$

Définition 2.8. — Le test de H_0 contre H_1 associé à la zone de rejet R est la stratégie suivante :

- Si $(x_1, \dots, x_n) \in R$, alors on rejette l'hypothèse H_0 et on accepte l'hypothèse alternative H_1 .
- Sinon, on accepte l'hypothèse H_0 .

L'erreur de première espèce d'un test avec zone de rejet R est si $\theta \in \Theta_0$, $\mathbb{P}_\theta((X_1, \dots, X_n) \in R)$.
L'erreur de seconde espèce d'un test avec zone de rejet R est si $\theta \in \Theta_1$, $1 - \mathbb{P}_\theta((X_1, \dots, X_n) \in R)$.

Définition 2.9. — *Le niveau d'un test de H_0 contre H_1 avec zone de rejet R est*

$$\alpha := \sup \{ \mathbb{P}_\theta((X_1, \dots, X_n) \in R), \theta \in \Theta_0 \}$$

La puissance de ce test est la fonction définie sur Θ_1 par

$$\beta(\theta) = \theta \rightarrow \mathbb{P}_\theta((X_1, \dots, X_n) \in R)$$

En pratique on considère le test suffisamment puissant si

$$\inf_{\Theta_1} \mathbb{P}_\theta((X_1, \dots, X_n) \in R) \geq 0.8$$

Remarque. — *Attention, le but d'un test est de rejeter H_0 . Si $R = \emptyset$, alors $\alpha = 0$, mais la puissance de ce test est 0. Il ne rejette pas assez. Il faut donc trouver le test de niveau prescrit $\alpha = 0.01$ par exemple, mais le plus puissant!*

3.3 Valeur p d'un test On teste une hypothèse simple $H_0 : \theta = \theta_0$ contre H_1 avec une zone de rejet $R = \{Z > z_\alpha\}$. Ici $Z = g(X_1, \dots, X_n)$ est une variable ne dépendant que des observations dont on connaît la loi sous P_{θ_0} . On a par exemple choisi z_α pour que le niveau du test soit $\alpha = \mathbb{P}_{\theta_0}(Z > z_\alpha)$.

Définition 2.10. — *On suppose que l'on a observé (x_1, \dots, x_n) et que $z_{obs} = g(x_1, \dots, x_n)$. La valeur- p (p value en anglais) est*

$$p := \mathbb{P}_{\theta_0}(Z > z_{obs}).$$

La valeur p représente combien peu probable est l'observation que l'on a faite. On rejette donc H_0 si $p < \alpha$, ce qui se produit si $z_{obs} > z_\alpha$ et donc $(x_1, \dots, x_n) \in R$, et on l'accepte sinon.

3.4 Construction d'un test à partir d'un intervalle de confiance
Supposons que l'on dispose d'un intervalle de confiance (construit à partir des observations) pour le paramètre θ de niveau $1 - \alpha$ pour le paramètre θ et que l'on désire tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. On décide alors de rejeter H_0 si θ_0 n'est pas dans l'intervalle de confiance. C'est un test de niveau inférieur à α . En revanche, on ne connaît pas sa puissance.

3.5 Exemple d'un test binomial Pour tester si un étudiant a travaillé on lui fait passer un QCM de $n = 20$ questions. S'il a travaillé il répond juste à chaque question avec la probabilité $p_1 = 0.6$, s'il n'a pas travaillé avec la probabilité $p_0 = 1/3$.

Le nombre de bonnes réponses est $X \sim \mathcal{B}(n, p)$.

On décide d'une zone de rejet de la forme $R = [k_c, 20]$ pour tester l'hypothèse $p = p_0$ contre $p = p_1$.

L'erreur de première espèce $\alpha = \mathbb{P}_{p_0}(X \geq k_c)$ est la probabilité de rejeter H_0 , ie de considérer que l'étudiant a travaillé, alors qu'il n'a pas travaillé. L'erreur de seconde espèce $\beta = \mathbb{P}_{p_1}(X < k_c)$ est la probabilité d'accepter H_0 , alors que H_1 est vraie, i.e. de penser que l'étudiant n'a pas travaillé alors qu'il a travaillé. Ici la puissance est $1 - \beta = \mathbb{P}_{p_1}(X \geq k_c)$.

k_c	$\alpha \times 100$	$\beta \times 100$
10	9	87
11	3,7	75
12	1	60
14	0,1	25

Quand k_c augmente, le niveau diminue, mais la puissance également.

Un test est un compromis niveau puissance.

Comment avoir α aussi proche de 5% et $\beta \geq 80\%$?

En augmentant le nombre de questions.

Si $n = 40$ et que l'on observe $k_{obs} = 20$, alors la p valeur est $\mathbb{P}(\mathcal{B}(40, p_0) \geq k_{obs}) = 0.02$ alors on rejette H_0 à 5 % et on l'accepte à 2 %. En fait on obtient un test de niveau proche de 5 % en prenant $k_c = 19$ et alors $\alpha = 4.4\%$ et $\beta = 96\%$;

3.6 Test du signe On considère un échantillon de sujets placés dans deux conditions différentes.

Par exemple un groupe de patients, qui la première fois ne reçoit pas de traitement et la seconde fois reçoit le traitement, et on mesure un indicateur : une variable ordinaire, c'est à dire dans un ensemble où on peut classer les valeurs, par exemple on mesure le taux de sucre dans le sang.

On dispose donc d'un vecteur de couples d'observations $((x_1, y_1), \dots, (x_n, y_n))$ et on suppose les variables $(X_i, Y_i)_{i \geq 1}$ IID.

L'hypothèse nulle est H_0 : les différences entre x_i et y_i sont dues au hasard. L'hypothèse alternative H_1 les différences ne sont pas dues au hasard (mais au traitement dans l'exemple précédent).

C'est un test non paramétrique.

On calcule les signes des $Y_i - X_i$ et on note D_+ le nombre de différences strictement positives, D_- le nombre de différences strictement négatives. On ne compte pas le nombre de différences nulles (hypothèse vérifiée dans le cas

de va à densité, on considère que la mesure n'est pas assez précise pour avoir détecté une différence), et donc on pose $N = D_+ + D_-$.

Sous H_0 on a $\mathbb{P}(Y_i > X_i) = \frac{1}{2}$ et la variable $D_+ \sim \mathcal{B}(N, 0.5)$ suit une loi binomiale.

On suppose que le traitement fait diminuer le taux de sucre s'il fonctionne, donc pour un niveau α la zone de rejet est $R = \{D_+ \leq k_\alpha\}$ avec k_α le plus grand entier k tel que $\mathbb{P}(S \leq k) < \alpha$ avec $S \sim \mathcal{B}(N, 0.5)$.

Exemple: On observe 14 sujets dans deux conditions, 2 différences positives, 10 négatives, 2 nulles. On a donc $N = 12$ et $D_{+,obs} = 2$. Soit $S \sim \mathcal{B}(12, 0.5)$, on a la valeur p

$$p = \mathbb{P}(S \leq D_{+,obs}) = 0.019$$

donc on rejette H_0 pour tout seuil α supérieur à 2%.