Une année d'agrégation

Jean Vereecke

20 novembre 2023

Table des matières

1 Probabilités et statistiques **3**
1.1 Définition d'un espace probabilisé . . . 3

2 Programme spécifique de l'option A **4**
2.1 Chaînes de Markov . . . 4

3 Généralités sur les groupes, groupes finis, groupes symétrique **11**
3.1 Rappels . . . 11
3.2 Générateurs d'un groupe . . . 12
3.3 Sous-groupes distingués . . . 12
3.4 Centre et commutateurs . . . 12
3.5 Opération d'un groupe sur un ensemble . . . 12

4 Cours de mercredi 18 octobre matin 1 – Matthieu Romagny **13**
4.1 Noyaux itérés et décomposition de Fitting . . . 13
4.2 Sous-espaces cycliques . . . 13
4.3 Réduction de Jordan . . . 14

5 20/10/2023 - Théorème d'Ascoli et applications – Arnaud Debussche **16**
5.1 Le cas métrique . . . 16

6 MOPS Probabilités et statistiques 24/10/2023 - Estimation statistique – Jean-Christophe Breton **19**

7 Cours Mercredi après-midi 25/10/2023 - Matrices sur les corps finis – Matthieu Romagny **22**
7.1 $GL_n(\mathbb{F}_q)$ et ses proches amis . . . 22
7.2 Matrices nilpotentes . . . 22
7.3 Matrices diagonalisables et trigonalisables . . . 23
7.4 Générateurs de GL_n : dilatations et transvections . . . 23

8 MOPS Probabilités et statistiques 26/10/2023 - Chaînes de Markov à espace d'états dénombrable – Nathalie Krell **25**
8.1 Introduction . . . 25
8.2 Loi d'une chaîne de Markov . . . 25
8.3 Récurrence et transience . . . 26
8.4 Temps de retour et nombre de visite . . . 26

9 Cours Mercredi après-midi 2 08/11/2023 - Formule du changement de variables – Jürgen Angst **27**

10 09/11/2023 - Gauss – Jean-Christophe Breton **29**

11 13/11/2023 - Géométrie Affine – Jérémy Le Borgne **31**

11.1 Espaces affines . . . 31
11.1.1 Définitions . . . 31
11.1.2 Sous-espaces affines . . . 32
11.1.3 Repères affines . . . 33
11.2 Applications affines . . . 33
11.2.1 Applications affines et repères affines . . . 34
11.2.2 Groupes affines . . . 34
11.2.3 Sous-groupe $HT(\mathcal{E})$ . . . 35
11.3 Barycentres . . . 36
11.3.1 Repère et coordonnées barycentriques . . . 36
11.3.2 Théorème fondamental de la géométrie affine . . . 37

12 16/11/2023 - Espérance conditionnelle – Jürgen Angst **38**

13 20/11/2023 - Géométrie projective – Jérémy Le Borgne **39**

Chapitre 1

Probabilités et statistiques

1.1 Définition d'un espace probabilisé

Probabilités conditionnelles

On rappelle la notion de probabilités conditionnelles.

Définition 1.1.1 (Probabilité conditionnelle)**.** Soient A et B deux évènements avec $\mathbb{P}(B) > 0$. On appelle **probabilité conditionnelle de A sachant B** la quantité

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{B}} = \mathbb{P}_B(A).$$

Cette notion permet de calculer plus simplement certaines probabilités à l'aide des deux formules suivantes.

Proposition 1.1.2 (Formule de Bayes)**.** Soient A et B deux évènements de probabilités non nulles. La formule de Bayes s'écrit alors

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

Formule des probabilité totales

Théorème 1.1.3 (Formule des probabilités totales)**.** Soit $(A_i)_{i\in I}$ un système complet d'évènements (*ie* des évènements deux à deux disjoints de probabilités non nulles avec I dénombrable et $\sum_{i\in I} \mathbb{P}(A_i) = 1$).

Alors pour tout évènement A, on a

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|A_i)\mathbb{P}(A_i).$$

Chapitre 2

Programme spécifique de l'option A

2.1 Chaînes de Markov

Chaînes de Markov à espace d'états fini ou dénombrable

Définition 2.1.1 (Processus stochastique)**.** Un **processus stochastique** est une famille $X = (X_t)_{t\in T}$ de variables aléatoires X_t. Si T est au plus dénombrable (respectivement indénombrable), le processus est alors dit à **temps discret** (respectivement à **temps continu**).

Chaînes de Markov à espace d'états fini

Soit E un ensemble fini (on va considérer que $E = [\![1, d]\!]$) et soit X_0 un état initial.

Définition 2.1.2. Le processus $(X_n)_{n\geq 0}$ est appelée une **chaîne de Markov homogène** ssi pour tout $y, x_0, \ldots, x_n \in E$ pour $n \geq 0$, les probabilités suivantes ne dépendent que de x_n et y,

$$\mathbb{P}(X_{n+1} = y | X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = y | X_n = x_n) = \mathbb{P}(X_1 = y | X_0 = x_n).$$

On note alors $P(x, y)$ (ou $P_{x,y}$) cette quantité que l'on appelera **probabilité de transition** de l'état x à y.

Remarque 2.1.3. La probabilité de transition $P(x, y)$ représente la probabilité de passer de l'état x à l'état y en un coup et, de ce fait, on a pour tout $x, y \in E$,

$$0 \leq P(x, y) \leq 1 \text{ et } \sum_{y\in E} P(x, y) = 1. \tag{2.1}$$

Une matrice vérifiant (2.1) sera appelé **matrice stochastique** et $P = (P(x, y))_{x,y\in E}$ est appelée matrice de transition.

Une chaîne de Markov permet de modéliser un phénomène dont le futur ne dépend du passé qu'à travers le présent.

Exemple 2.1.4 (Prédiction de météo)**.** On cherche à prédire la météo par une chaîne de Markov sur l'ensemble d'états $\{0, 1\}$ où 0 modélise le fait qu'il pleuve et 1 qu'il ne pleuve pas. On note X_n alors le temps du n-ème, on considère alors que c'est une chaîne de Markov de matrice de transition $P = \begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix}$ où $p, q \in [0, 1]$.

Exemple 2.1.5 (Marche aléatoire sur un cube)**.** On considère la marche aléatoire sur un cube $E = \{0, 1\}^d$ avec la matrice de transition $P(x, y) = \frac{1}{d}\delta_{|x-y|,1}$.

Exemple 2.1.6 (Urnes d'Ehrenfest)**.** On a d boules numérotées de 1 à d, on a alors une matrice de transition P avec $P(x, y) = \frac{x}{d}$ si $y = x - 1$, $P(x, y) = \frac{d-x}{d}$ si $y = x + 1$ et $P(x, y) = 0$ sinon.

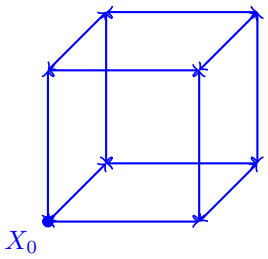

FIGURE 2.1 – Illustration de la marche aléatoire du cube

Proposition 2.1.7. Soit X une chaîne de Markov sur un espace de probabilité de matrice de transition P et de loi initiale μ_0. Alors pour $n \geq 0$ et $x_0, \ldots, x_n \in E$, on a

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \mu_0(x_0) \prod_{k=0}^{n-1} P(x_k, x_{k+1}). \tag{2.2}$$

Exemple 2.1.8 (Retour sur la prédiction de météo)**.** Avec l'exemple précedent, on prend $\mu_0 = \mathbb{P}_{X_0}$, la probabilité qu'il ne pleuve pas pendant 3 jours de suite est

$$\mathbb{P}(X_0 = X_1 = X_2 = 1) = \mu_0(1)(1-q)^2.$$

Sachant qu'il ne pleut pas aujourd'hui, la probabilité qu'il ne pleuve pas les deux prochains jours est

$$\mathbb{P}(X_1 = X_2 = 1 | X_0 = 1) = (1-q)^2.$$

Propriétés de Markov faible et forte

Espace d'états fini

Soit X une chaîne de Markov sur E. X peut-être vu comme une variable aléatoire à valeurs dans $\Pi = E^{\mathbb{N}}$ muni des tribus cylindriques $(\mathcal{B}_n(\Pi))_{n\geq 0}$ où $\mathcal{B}_n(\Pi) = \sigma(X_0, \ldots, X_n)$.
Si la loi de X_0 est μ_0, on note $\mathbb{P}_{\mu_0}$ la loi de X et si $\mu_0 = \delta_x$, on note $\mathbb{P}_{\delta_x} = \mathbb{P}_x$.

Proposition 2.1.9. La loi d'une chaîne de Markov sur E de matrice de transition P et de loi initiale μ_0 est l'unique probabilité sur $(\Pi, \mathcal{B}(\Pi))$ caractérisée par

$$\mathbb{P}(\{x \in \Pi : x_0 = y_0, \ldots, x_n = y_n\}) = \mu_0(y_0)P(y_0, y_1) \cdots P(y_{n-1}, y_n).$$

On introduit le processus décalé de n $X_{n+} = (X_{n+k})_{k\geq 0}$.

Théorème 2.1.10 (Propriété de Markov faible)**.** Soit X une chaîne de Markov sur E de loi initiale μ_0. Pour tout évènement $A \in \mathcal{B}(\Pi)$ et $n \in \mathbb{N}$, $\mathbb{P}(X_{n+} \in A | X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}_{x_n}(A)$.

Les équations de Chapman-Kolmogorov

On note μ_n la loi de X_n. Pour $y \in E$, $\mu_{n+1}(y) = \cdots = \sum_{x\in E} \mu_n(x)P(x, y)$.

Soit h une fonction mesurable bornée de E dans $\mathbb{R}$, on définit Ph la fonction mesurable bornée de E dans $\mathbb{R}$ telle que pour tout $x \in E$, $Ph(x) = \sum_{y\in E} P(x, y)h(y)$.

Soit μ une mesure positive (respectivement de probabilité), alors la mesure positive (respectivement de probabilité) μP est défini par pour tout $y \in E$, $\mu P(y) = \sum_{x\in E} \mu(x)P(x, y)$, d'où $\mu_{n+1} = \mu_n P$.

Remarque 2.1.11. Pour tout $n \geq 0$, et tous $x, y \in E$, $P^n(x,y) = \mathbb{P}(X_{n+k} = y | X_k = x)$ pour tout $k \geq 0$.

Proposition 2.1.12. On a pour tous $n, m \geq 0$, pour tous $x, y \in E$, $P^{n+m}(x,y) = \sum_{z \in E} P^n(x,z) P^m(z,y)$.

En conséquence, la loi à l'instant n du processus X ne dépend que de la loi μ_0 de X_0 et de sa matrice de transition P, elle vaut $\mu_n = \mu_0 P^n$. Pour toute h fonction de E dans $\mathbb{R}$,

$$\mathbb{E}\left[h(X_n) | X_0 = x\right] = P^n h(x) \quad \text{et} \quad \mathbb{E}[h(X_n)] = \mu_0 P^n h.$$

Exemple 2.1.13 (Retour sur la prédiction de météo)**.** Si on prend $p = 0,7$ et $q = 0,4$. Sachant qu'il pleut aujourd'hui, on souhaite connaître la probabilité qu'il pleuve dans 4 jours, elle vaut $\mathbb{P}(X_4 = 0 | X_0 = 0) = 0,5749$. On suppose qu'il pleuve le premier jour avec probabilité $p_0 = 0,1$, on souhaite connaître la probabilité qu'il pleuve dans 4 jours, elle vaut $\mathbb{P}(X_4 = 0 | X_0 = 0) = 0,5876$.

Classification des états, transience, récurrence positive ou nulle

Définition 2.1.14. Notons T_x le premier temps d'atteinte de x à partir de la première transition ie

$$T_x = \inf\{n \geq 1 : X_n = x\}.$$

On note $f_x = \mathbb{P}_x(T_x < +\infty)$ la probabilité que le processus revienne au moins une fois en x partant de x, alors

$$f_x = \mathbb{P}(\exists n \geq 1, X_n = x | X_0 = x) = \sum_{n \geq 1} \mathbb{P}_x(T_x = n).$$

Remarque 2.1.15. $\{T_x = n\} = \{X_1 \neq x, \ldots, X_{n-1} \neq x, X_n = x\}$.

Définition 2.1.16. Si $\mathbb{P}_x(T_x < \infty) = 1$, on dit alors que q est un **état récurrent**. Sinon, on dit qu'il est **transient** (ou **transitoire**).
Si $P(x,x) = 1$, on dit que x est **absorbant**.

Remarque 2.1.17. Si x est un état récurrent, partant de x, on est certain de retourner en x et aussi de suite, on peut ainsi une infinité de fois en x.

En revanche, si x est transient, on a une probabilité non-nulle de ne jamais y revenir. Partant de x, le nombre de fois N_x que le processus passe en x suit une géométrique de paramètre $1 - f_x$ donc le nombre moyen de passage en x est $\frac{1}{1-f_x}$.

Proposition 2.1.18. Soit $x \in E$, alors x est **récurrent** (resp. **transient**) si et seulement si

$$\sum_{n \geq 0} P^n(x,x) = +\infty \quad \left(\text{resp. } \sum_{n \geq 0} P^n(x,x) < +\infty \right).$$

Démonstration. $\mathbb{E}_x(N_x) = \mathbb{E}_x\left(\sum_{n=0}^{+\infty} \mathbb{1}_{X_n = x}\right) = \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = x) = \sum_{n=0}^{+\infty} P^n_{x,x}$. □

Remarque 2.1.19. Si l'état x est transient, $\lim_{n \to +\infty} P^n(x,x) = 0$.

Un critère analytique de récurrence

Soit $x \in E$ et soit $t \in [0,1]$, on pose

$$U(x,t) = \mathbb{E}_x\left[t^{T_x} \mathbb{1}_{\{T_x < +\infty\}}\right] = \sum_{k \geq 1} \mathbb{P}_x(T_x = k) t^k,$$

alors $U(x, \cdot)$ est la fonction génératrice du temps de retour en x.

Par définition $U(x,1) = \mathbb{P}_x(T_x < +\infty)$ et $U'(x,1) = \mathbb{E}_x[T_x \mathbb{1}_{\{T_x<+\infty\}}]$.
On pose alors

$$G(x,t) = \mathbb{E}_x\left[\sum_{k\geq 0} \mathbb{1}_{X_k=x} t^k\right] = \sum_{k\geq 0} P^k(x,x)t^k$$

et donc $G(x,1) = \mathbb{E}_x(N_x)$.

Théorème 2.1.20. Pour tous $x \in E$ et $t \in [0,1]$, on a $G(x,t) = \frac{1}{1-U(x,t)}$. En particulier, l'état x est récurrent si et seulement si $G(x,1) = \sum_{k\geq 1} P^k(x,x) = +\infty$

Démonstration. On a

$$\sum_{k\geq 0} \mathbb{1}_{\{X_k^x=x\}} t^k = 1 + \sum_{n\geq 1} t^{T_x^n} \mathbb{1}_{\{T_x^n<+\infty\}}.$$

Alors

$$T_x^n = \sum_{k=1}^{n} (T_x^k - T_x^{k-1}) \text{ avec } T_x^0 = 0$$

et

$$\mathbb{E}_x\left[t^{T_x^n}\mathbb{1}_{\{T_x^n<+\infty\}}\right] = \mathbb{E}_x\left[\prod_{k=1}^{n} t^{T_x^k - T_x^{k-1}} \mathbb{1}_{\{T_x^k - T_x^{k-1}<+\infty\}}\right] = \mathbb{E}_x\left[t^{T_x}\mathbb{1}_{\{T_x<+\infty\}}\right]^n = U(x,t)^n.$$

Donc

$$G(x,t) = 1 + \sum_{n\geq 1} U(x,t)^n = \frac{1}{1-U(x,t)}.$$

□

Communication entre états et irréductibilité

On notera $x \rightsquigarrow y$ si et seulement si x **mène** en y ie il existe $n \in \mathbb{N}$ tel que $P_{x,y}^n > 0$. On dira que x et y **communiquent** (noté $x \sim y$) si et seulement si $x \rightsquigarrow y$ et $y \rightsquigarrow x$. Il s'agit d'une relation d'équivalence ce qui permet de décomposer l'espace d'états E en différentes classes à l'intérieur de chacune les états communiquent entre eux.

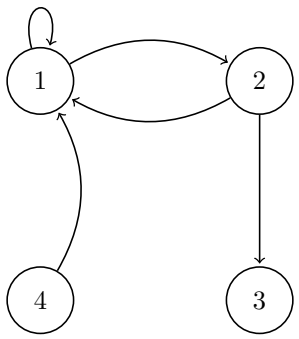

FIGURE 2.2 – Exemple de chaînes de Markov

Dans l'exemple donné en figure 2.2, $1 \sim 2$, $2 \rightsquigarrow 3$ et $4 \rightsquigarrow 1$.

Définition 2.1.21. Une chaîne de Markov est dite **irréductible** si et seulement si elle ne possède qu'une seule classe pour la relation d'équivalence précédente ie que tous les états communiquent entre eux.

Exemples 2.1.22. Pour $E = \{0,1,2\}$ et $P = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$, on a une chaîne de Markov irréductible.

Pour $E = \{0,1,2,3\}$ et $P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3/4 & 1/4 & 0 & 0 \\ 1/8 & 0 & 7/8 & 0 \\ 0 & 0 & 1/9 & 8/9 \end{pmatrix}$, on a une chaîne de Markov non irréductible (en réduisant à $\{0,1\}$ elle est irréductible).

Remarque 2.1.23. Soient $x, y \in E$ avec $x \sim y$ et x récurrent alors y aussi. Par conséquent, tous les éléments de la classe sont de même nature.

Théorème 2.1.24. Si $|E| < +\infty$, il existe au moins un état récurrent. Si de plus la chaîne est irréductible, alors tous les états sont récurrents.

Mesure stationnaire (existence et unicité)

Définition 2.1.25. Une mesure π sur E est dite **invariante** pour P, si $\pi P = \pi$. Si de plus π est une mesure de probabilité, on dit que π est une **probabilité invariante** (ou **stationnaire**).

Remarque 2.1.26. En cardinal fini, il y a toujours l'existence d'une probabilité invariante mais pas en cardinal infini. Par exemple, il n'existe pas de probabilité invariante à la marche aléatoire symétrique sur $\mathbb{Z}$, mais la mesure uniforme sur $\mathbb{Z}$ est invariante.

Définition 2.1.27. Considérons $x \in E$ et $R(x) = \{n \geq 0 : P^n_{x,x} > 0\}$. On définit alors la **période** $p(x)$ d'un état x comme le plus grand commun diviseur de $R(x)$ ie $p(x) = \text{pgcd}(R(x))$.

L'état x est dit **apériodique** si et seulement si $p(x) = 1$.

[exemple chaîne de markov 1->1 1->2 2->1 2->3 3->1]

Remarque 2.1.28. La péridocité est une propriété de classe.

Exemple 2.1.29. L'rrne d'Ehrenfest et la marche aléatoire ont une période de 2.

Théorème 2.1.30. Soit E un espace d'états fini. Si P est irréductible et apériodique alors il existe une unique probabilité invariante π tel que pour tout μ_0 loi initiale, X converge en loi vers π ie $\mu_0 P^n \xrightarrow[n\to+\infty]{} \pi$.

Remarque 2.1.31. $\pi P = \pi$ donc si $X_n \sim \pi$ alors les X_m pour $m \geq n$ aussi. π est une probabilité stationnaire et invariante.

Dans le cas $|E| < +\infty$, il existe toujours une probabilité stationnaire, l'irréductibilité donne l'unicité. La marche aléatoire sur le cube vérifie $P^{2n+1} = P \neq P^2 = P^{2n}$.

Proposition 2.1.32. Si P est irréductible alors P admet au moins une probabilité invariante, si π est une telle probabilité alors $\pi(x) > 0$ pour tout $x \in E$.

Théorème 2.1.33. Soit P est irréductible, les assertions suivantes sont équivalentes :

1. il existe une unique invariante π ;
2. la mesure π définie par : pour $x \in E$, $\pi(x) = \frac{1}{\mathbb{E}(T_x)}$ est la probabilité invariante ;
3. tous les états sont récurrents positifs ;
4. il existe un état récurrent positif.

Théorèmes de convergence : loi des grands nombres, apériodicité et convergence en loi

Théorème 2.1.34 (Théorème ergodique)**.** Soit P irréductible, alors pour toute mesure initiale ν, on a la convergence $\mathbb{P}_\nu$-presque-sûrement

$$\frac{1}{n}\sum \mathbb{1}_{\{X_k = x\}} \xrightarrow[n\to+\infty]{} \frac{1}{\mathbb{E}(T_x)}.$$

Si de plus la chaîne est récurrente positive alors pour tout f mesurable bornée sur E, on a la convergence $\mathbb{P}_\nu$-presque-sûrement

$$\frac{1}{n}\sum f(X_k) \xrightarrow[n\to+\infty]{} \int_E f \mathrm{d}\pi.$$

Théorème 2.1.35. Soit X une chaîne de Markov irréductible récurrente positive de probabilité invariante π,

1. si la chaîne est apériodique, pour tous $x, y \in E$, $\lim_{n\to+\infty} P^n(x,y) = \pi(y)$ et pour tout μ loi initiale, $\lim_{n_to+\infty} \mathbb{P}(X_n = y) = \pi(y)$;
2. si la chaîne est périodique de période d, pour tous $x, y \in E$, il existe $0 \leq n < d$ tel que si $P^n(x,y) \neq 0$, $n = r[d]$ et $\lim_{n_to+\infty} \mathbb{P}^{nd+r}(x,y) = d\pi(y)$.

Exemple de la marche aléatoire simple

Théorème 2.1.36 (Marche aléatoire sur $\mathbb{Z}$)**.** La marche aléatoire sur $\mathbb{Z}$ est irréductible, est récurrente nulle dans le cadre de la marche aléatoire symétrique et transiente dans le cadre de la marche aléatoire asymétrique.

Démonstration. La chaîne est irréductible donc tous les états sont de même nature, on regarde donc $x = 0$. Pour $n \in \mathbb{N}$, si $n = 2k+1$, $P^n(0,0) = 0$, sinon $n = 2k$, $P^n(0,0) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$ et grâce à la formule de Stirling,

$$P^n(0,0) \underset{k\to+\infty}{\sim} \frac{(4p(1-p))^k}{\sqrt{\pi k}}$$

qui est le terme général d'une série convergente si et seulement si $2p \neq 1$. Et

$$\mathbb{E}_x(T_x | T_x < +\infty) = \frac{U'(x,1)}{U(x,1)} = \frac{1}{1 - \frac{1}{2(p\vee(1-p))}}$$

ce qui donne le résultat. □

Théorème 2.1.37 (Polya 1921)**.** Soit X la marche aléatoire simple sur $\mathbb{Z}^d$ avec pour suite d'incréments U. Pour $d \leq 2$, la marche aléatoire X est récurrente, pour $d \geq 3$, elle est transciente.

Démonstration. Si $d = 1$, on a déjà le résultat.
Si $d = 2$, on note $X = (X^1, X^2)$ et $U = (U^1, U^2)$. Les variables $S_n = U_n^1 + U_n^2$ et $D_n = U_n^1 - U_n^2$ sont indépendantes et de même loi, $\mathbb{P}(S_n = +1) = \mathbb{P}(S_n = -1) = \frac{1}{2}$. On se ramène au cas de marche aléatoire sur $\mathbb{Z}$ et

$$P^{2k}(0,0) = \mathbb{P}(X_{2k}^1 + X_{2k}^2 = X_{2k}^1 - X_{2k}^2 = 0 | X_0 = 0) = \mathbb{P}(S_1 + \cdots + S_{2k} = 0)\mathbb{P}(D_1 + \cdot + D_{2k} = 0) \underset{k\to+\infty}{\sim} \frac{1}{\pi k}.$$

Si $d \geq 3$, si on admet que $P^n(0,0) \sim \frac{\text{cst}}{k^{d/2}}$, c'est le terme général d'une série convergente ce qui donne le résultat.
[Norris d=3 est traité]
Soit Φ la fonction caractéristique des U_n,

$$\Phi(t) = [\exp(i\langle t, U_i\rangle)] = \frac{\cos t_1 + \cdots + \cos t_d}{d},$$

par Fubini, $\frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \Phi_{X_n}(t)\mathrm{d}t = \mathbb{P}(X_n = 0)$ d'où $\sum_{k\geq 0} P^{2k}(0,0) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{\mathrm{d}t}{1-\Phi^2(t)}$ qui diverge. □

Exemple 2.1.38 (Marche aléatoire réfléchie sur $\mathbb{N}$)**.** On a le processus X de probabilité de transition P, $P(0,0) = 1 - P(0,1) = a \in [0,1[$ et $P(x,x+1) = 1 - P(x,x-1) = p \in]0,1[$ pour $x \geq 1$. On peut montrer que

$$U(0,t) = ta + \frac{1-a}{2p}(1 - \sqrt{1 - 4t^2p(1-p)})$$

et que la marche est transiente (resp. récurrent nul, récurrent positive) ssi $p > 1/2$ (resp. $p = 1/2$, $p < 1/2$).

Chaîne de Markov et méthode de monte Carlo : algorithme de Métropolis

On a E de cardinal fini mais grand, on considère π probabilité sur E

Exemple 2.1.39 (Modèle d'Ising)**.**

On veut générer une chaîne de Markov qui converge vers π, on se donne une matrice de transition Q appelée matrice de sélection telle que pour tous $x, y \in E$, si $Q(x,y) > 0$, $Q(y,x) > 0$.

Pour $x \neq y$, posons $R(x,y) = \min\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}, 1\right)$ si $Q(x,y) \neq 0$ et vaut 0 sinon.

On construit alors une matrice de transition P telle que $P(x,y) = Q(x,y)R(x,y)$ si $x \neq y$ et $P(x,y) = 1 - \sum_{y \neq x} P(x,y)$.

Proposition 2.1.40. Si π charge tous les points de E, alors la matrice P est réversible par rapport à π. De plus, elle est apériodique et irréductible ssi Q est irréductible.

Démonstration. □

[Algorithme de Métropolis]

Remarque 2.1.41. Si Q est symétrique, il suffit de tester $\pi(y) < \pi(x)$.

Chapitre 3

Généralités sur les groupes, groupes finis, groupes symétrique

Références : [Per96].

3.1 Rappels

On suppose connues les notions de groupe, groupe abélien, de sous-groupe, d'(homo)morphisme de groupes.

Nous noterons généralement multiplicativement les lois de groupe, l'élément neutre sera alors noté 1, l'inverse de g sera noté g^{-1}. Cette règle aura une exception pour les groupes abéliens qui seront notés additivement.

Le cardinal d'un groupe fini est également appelé son **ordre**. Si p est un nombre premier, on appelle p-groupe un groupe dont le cardinal est une puissance de p. Si $g \in G$, l'ordre de g est le plus petit entier $n > 0$ (s'il existe) tel que $g^n = 1$. C'est aussi l'ordre du groupe $\langle g \rangle$.

Le noyau d'un morphisme $f \in \mathbf{Hom}(F, G)$ est le sous-groupe de G défini par :

$$\ker(f) = \{g \in G : f(g) = 1\}.$$

L'image de f est un sous-groupe de H, noté $\mathrm{Im}(f)$. Un isomorphisme est un morphisme de groupes bijectif. Un **automorphisme** d'un groupe G est un isomorphisme de G sur G. Un exemple d'automorphisme est fourni par les **automorphismes intérieurs**. Un tel automorphisme i_g est donné, pour $g \in G$, par la formule $i_G(x) = gxg^{-1}$.

Si $H < G$, on appelle **classe à gauche** de l'élément $a \in G$ relativement à H le sous-ensemble aH et on définit de même les classes à droite Ha. Les classes à gauche forment une partition de G. Leur ensemble est noté G/H. Ce n'est pas un groupe en général. Le cardinal de G/H est appelé l'**indice** de H dans G et est noté $(G : H)$. Lorsque le groupe est fini, la considération des classes à gauche conduit au théorème suivant :

Théorème 3.1.1 (Lagrange)**.** Si H est un sous-groupe du groupe fini G, l'ordre de H et l'indice de H dans G divisent l'ordre de G. Précisément, on a :

$$|G| = |H||G/H| = |H|(G : H).$$

En particulier, l'ordre d'un élément $g \in G$ divise l'ordre de G.

Le groupe des bijections (ou permutations) d'un ensemble E s'appelle le **groupe symétrique** de E et est noté $\mathfrak{S}(E)$. Si E et E' ont le même cardinal alors $\mathfrak{S}(E) \simeq \mathfrak{S}(E')$. Lorsque $E = [\![1, n]\!]$, on parle du groupe symétrique standard. Le cardinal de ce groupe est $n!$.

Le groupe symétrique contient des permutations remarquables : les **cycles** d'ordre k. Un tel cycle est noté $\sigma = (a_1, \ldots, a_k)$ avec les $a_i \in E$ deux à deux distincts et la notation signifie que l'on a $\sigma(a) = a$ si a n'est pas l'un des a_i et $\sigma(a_i) = \sigma(a_{i+1})$ (où l'indice est pris modulo k). Un tel cycle est un élément d'ordre k. Pour $k = 2$, on parle de **transpositions**.

Le groupe $\mathfrak{S}_n$ est muni d'un morphisme surjectif, appelé signature, et noté $\varepsilon : \mathfrak{S}_n \to \{+1, -1\}$ que l'on peut définir de multiples façons mais dont nous retiendrons les propriétés suivantes :

1. si τ est une transposition, on a $\varepsilon(\tau) = -1$,
2. plus généralement, si σ est un cycle d'ordre k, on a $\varepsilon(\sigma) = (-1)^k$.

Le noyau de ε est formé des permutations paires. C'est un groupe de cardinal $\frac{n!}{2}$ appelé **groupe alterné** et noté $\mathfrak{A}_n$.

Enfin, le lecteur est supposé avoir une certaine familiarité avec quelques objets élémentaires comme les groupes additifs $\mathbb{Z}/n\mathbb{Z}$ des congruences modulo n ou le groupe de Klein $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$,...

3.2 Générateurs d'un groupe

Proposition-Définition 3.2.1. Soient G un groupe et $A \in \mathcal{P}(G)$. Il existe un plus petit-sous groupe H de G contenant A. On dit que H est le sous-groupe engendré par A, ou que les éléments de A sont des **générateurs** de H. On note $H = \langle A \rangle$.

Démonstration. L'existence de H peut se voir de deux manières :
- par l'extérieur : on considère tous les sous-groupes de G contenant A et leur intersection convient ;
- par l'intérieur : on suppose que A est non vide, on pose alors $A^{-1} = x^{-1}, x \in A$ puis $H = \{a_1 \cdots a_n, n \in \mathbb{N}, a_i \in A \cup A^{-1}\}$. Alors H est un groupe contenant A et convient.

□

Exemples 3.2.2. Voici deux exemples.

1. *Groupes monogènes et cycliques.*
   Un groupe G engendré par un élément a est dit monogène. Il est isomorphe à $\mathbb{Z}$ ou un $\mathbb{Z}/n\mathbb{Z}$. Dans le second cas, G est dit cyclique. En particulier, si $|G| = p$ est un nombre premier, G n'a pas de sous-groupes non trivial (en vertu du théorème de Lagrange), donc si $a \in G \setminus \{1\}$, $G = \langle a \rangle$, donc cyclique et on a $G \simeq \mathbb{Z}/p\mathbb{Z}$.
2. *Groupes symétrique $\mathfrak{S}_n$ et alterné $\mathfrak{A}_n$.*
   (a) Les transpositions engendrent $\mathfrak{S}_n$, on peut même se limiter aux transpositions $(1,2)$, $(1,3)$, ...,$(1,n)$ ou encore $(1,2)$, $(2,3)$, ...,$(n-1,n)$ comme on le voit aisément par récurrence sur n. Le lecteur montrera à titre d'exercice que la transposition $(1,2)$ et le n-cycle $(1,2,\ldots,n)$ engendrent $\mathfrak{S}_n$ (cf. proposition 3.5.1).
   (b) Les 3-cycles engendrent $\mathfrak{A}_n$ pour $n \geq 3$. En effet $\mathfrak{A}_n$ est engendré par les produits pairs de transpositions et on a les formules : $(a,b)(b,c) = (a,b,c)$ et $(a,b)(a,c) = (a,c,b)$ (ce qui prouve au passage que tous les cycles d'ordre 3 sont dans $\mathfrak{A}_n$), $(a,b)(c,d) = (a,b)(a,c)(a,c)(c,d) = (a,c,b)(a,b,c)$. Nous verrons, à propos des groupes classiques, de nombreux autres exemples de générateurs.

3.3 Sous-groupes distingués

Définition 3.3.1. Soient G un groupe et H un sous-groupe de G. On dit que H est **distingué** dans G ssi il est invariant par automorphisme intérieur ie on a pour tout $a \in G$ et tout $h \in H$, $aha^{-1} \in H$. On note alors $H \triangleleft G$.

Remarques 3.3.2.
1. La condition ci-dessus équivaut à dire que pour tout $a \in G$, on a $aH = Ha$, ie l'égalité des classes à droite et à gauche modulo H.
2. Si $f : G \to G'$ est un morphisme de groupes, son noyau $\ker(f)$ est un sous-groupe distingué de G.
3. Réciproquement, si on a $H \triangleleft G$, le quotient G/H, ensemble des classes à gauche (ou à droite) est muni d'une structure de groupe et on a un morphisme surjectif $p : G \to G/H$ de noyau H.
4. Enfin on définit une **suite exacte** : $1 \to N \xrightarrow{i} G \xrightarrow{p} H \to 1$. Dans cette écriture, N, G, H sont des groupes i, p des morphismes et la suite est dite exacte ssi i est injectif, p est surjectif et $\mathrm{Im}(i) = \ker(p)$. Lorsque les groupes spnt abéliens et notés additivement, on écrit les suites exactes avec des 0.

3.4 Centre et commutateurs

3.5 Opération d'un groupe sur un ensemble

Proposition 3.5.1 (Proposition 4.10)**.**

Chapitre 4

Cours de mercredi 18 octobre matin 1 – Matthieu Romagny

Références : [MM12]

4.1 Noyaux itérés et décomposition de Fitting

Lemme 4.1.1 (Lemme des noyaux itérés)**.** Soit k un corps et soit E un k-espace vectoriel de dimension finie. Soit $f \in \mathcal{L}(E)$.

1. La suite $(\ker(f^k))_{k\in\mathbb{N}}$ est une suite strictement croissante puis stationnaire en un entier $p \in \mathbb{N}$.
2. La suite $(\mathrm{Im}(f^k))_{k\in\mathbb{N}}$ est une suite strictement décroissante puis stationnaire en ce même entier p.
3. On a la propriété suivante : $E = \ker(f^p) \oplus \mathrm{Im}(f^p)$. Si on note $K = \ker(f^p)$ et $I = \mathrm{Im}(f^p)$, la propriété est caractérisé par la décomposition suivante : $f|_K$ est nilpotente et $f|_I$ est inversible.
4. La suite $\delta = (\dim(\ker f^k) - \dim(\ker f^{k+1}))_{k\in\mathbb{N}}$ décroît jusque 0 ie les $\ker(f^k)$ croissent de moins en moins vite.

Démonstration. Les deux premiers points sont simples.

Concernant le troisième point, par définition de K, $f|_K$ est nilpotent d'indice p. Par définition de I, $f|_I$ est inversible car surjective. De plus, il reste à montrer que $E = K \oplus I$ (exercice à faire).

Concernant le quatrième point, on considére la composée u

$$\ker(f^{k+1}) \xrightarrow{f} \ker(f^k) \xrightarrow{\text{quotient}} \ker(f^k)/\ker(f^{k-1}).$$

Le noyau de u est alors $\ker(f^k)$, on en déduit que

$$\ker(f^{k+1})/\ker(f^k) \xhookrightarrow{\overline{f}} \ker(f^k)/\ker(f^{k-1}).$$

□

4.2 Sous-espaces cycliques

Soit k un corps et soit E un k-espace vectoriel de dimension finie. Soit $f \in \mathcal{L}(E)$.

Définition 4.2.1.
- Un **vecteur cyclique** (pour f) est un vecteur $x \in E$ tel que les $f^k x$ engendrent E.
- Un **sous-espace cyclique** (pour f) est un sous-espace $F \subset E$ qui est f-stable et qui possède un vecteur cyclique ie F est de la forme $E_{f,x} := \mathrm{Vect}_k(x, fx, \ldots)$.
- Un **endomorphisme cyclique** est un $f \in \mathcal{L}(E)$ tel que E est cyclique pour f.

Remarques 4.2.2.
- A priori, rien ne garantit l'existence d'un vecteur cyclique.
- Pour $f = \mathrm{id}_E$, on a alors $E_{f,x} = Kx$.

Définition 4.2.3. Si $x \in E$, on appelle **polynôme minimal ponctuel** noté $\mu_{f,x}$ l'unique générateur unitaire de l'idéal noyau de $k[X] \ni P \mapsto P(f)x \in E$.

Remarques 4.2.4.
— L'image du morphisme considéré est l'espace $E_{f,x}$, on obtient donc un isomorphisme $k[X]/(\mu_{f,x}) \simeq E_{f,x}$ donc E est cyclique ssi $\deg(\mu_{f,x}) = n$.
— On a $\mu_{f,x} | \mu_f$ car $\mu_f(f)x = 0$.

Lemme 4.2.5. Il existe $x \in E$ tel que $\mu_{f,x} = \mu_f$.

Démonstration. On écrit $\mu_f = P_1^{\alpha_1} \cdots P_r^{\alpha_r}$, une décomposition en facteurs premiers avec des P_i deux à deux distincts. D'après le lemme des noyaux, $E = \bigoplus_{i=k}^{r} E_k$ avec pour tout $k \in [\![1, r]\!]$, $E_k = \ker P_k^{\alpha_k}(f)$. Ainsi sur E_k, $P_k(f)$ est nilpotent d'ordre α_k.
Soit $x_k \in \ker P_k^{\alpha_k}(f) \setminus \ker P_k^{\alpha_k - 1}(f)$, alors $P_k^{\alpha_k}(f)x_k = 0$ mais $P_k^{\alpha_k - 1}(f)x_k \neq 0$ donc $\mu_{f,x_k} = P_k^{\alpha_k}$.
Posons $x = x_1 + \cdots + x_r$ et vérifions qu'il convient. Soit $Q \in k[X]$ tel que $Q(f)x = 0$. Comme chaque E_k est $Q(f)$-stable, on a $Q(f)x_1 + \cdots + Q(f)x_r = 0$ donc (comme la somme est directe) chaque $Q(f)x_k$ sont nuls ie $P_k^{\alpha_k} | Q$. Les $P_k^{\alpha_k}$ étant premiers entre eux, on en déduit que $\mu_k | Q$ donc $\mu_{f,x} = \mu_f$. □

Application : $\deg(\mu_f) \leq \operatorname{rg}(f)$.

Démonstration. Pour x vérifiant $\mu_{f,x} = \mu_f$, on a $E_{f,x} = \operatorname{Vect}_k(x, fx, \ldots, f^{p-1}x)$ de dimension p où $p = \deg(\mu_f)$, ainsi $p - 1 \leq \operatorname{rg}(f)$ d'où le résultat.
L'inégalité est optimale car elle est atteinte pour 0 et les blocs de Jordan. Il s'agit d'une véritable inégalité pour id_E notamment. □

Lemme 4.2.6. Pour x vérifiant $\mu_{f,x} = \mu_f$, alors le sous-espace $E_{f,x}$ possède un supplémentaire stable par f.

Remarques 4.2.7.
— Les endomorphismes f tels que tout sous-espace stable possède un supplémentaire stables sont appelées semi-simples.
— Tout ça est bien sûr utilisable pour la leçon « sous-espace stables ».

Démonstration. Soit $p = \deg(\mu_f)$, on complète la famille libre $(x, fx, \cdots, f^{p-1}x)$ en une base de E en $(e_1, \ldots, e_n)$. On pose alors

$$F = \bigcap_{j \geq 0} \ker(e_p^* \circ f^j) = \{y \in E : \forall j \geq 0, e_p^*(f^j y) = 0\}.$$

On montre que F est f-stable, $F \cap E_{f,x} = \{0\}$ et pour conclure, on calcule la dimension de F. Comme les f^j pour $j \geq p$ est combinaison linéaire des précédents, on a $F = \bigcap_{j=0}^{p-1} \ker(e_p^* \circ f^j)$, on observe que ces p formes linéaires sont bien indépendentes.
Comme par ailleurs, $\sum_{j=0}^{p-1} a_j f^j x \in E_{f,x}$, $\sum_{j=0}^{p-1} a_j f^j = 0$, donc pour tout j, $a_j = 0$ car $(x, fx, \cdots, f^{p-1})$ est libre.
L'intersection des noyaux de p formes linéaires indépendentes est de dimension $n - p$ (le dual de $\bigcap_{i=0}^{p-1} \ker(\varphi_i)$ est $\operatorname{Vect}_k(\varphi_1, \ldots, \varphi_p)$. Finalement, $\dim F = n - p$ et $\dim E_{f,x} = p$ donc $F \oplus E_{f,x} = E$. □

Remarque 4.2.8. Ne pas hésiter à y penser pour la leçon sur les formes linéaires : il s'agit d'un énoncé dans lequel a priori aucune forme linéaire n'apparaît qui nécessite celles-ci !

4.3 Réduction de Jordan

Définition 4.3.1 (Bloc de Jordan)**.** On appelle **bloc de Jordan nilpotent** d'ordre n la matrice $J_n \in \mathcal{M}_n(k)$ ayant des 1 au dessus de la diagonale et des 0 partout ailleurs et bloc de Jordan de paramètre $\lambda \in K$ et d'échelon n la matrice $J_n(\lambda) = J_n + \lambda I_n$.

Remarque 4.3.2. Pour tout $n, k \in \mathbb{N}$, J_n^k est nilpotente de rang $(m - k)_+$.

Nous allons traiter le cas nilpotent.

Proposition 4.3.3 (Unicité de la réduction de Jordan dans le cas nilpotent)**.** Soit $f \in \mathcal{L}(E)$ tel qu'il existe $\deg(\mu_f) = d_1 \geq \cdots \geq d_r$ et $\mathcal{B}$ base de E dans laquelle $\mathcal{M}_{\mathcal{B}}(f)$ est la matrice diagonale par blocs, les blocs étant $J_{d_1}, \ldots, J_{d_r}$. Alors pour tout k, $|\{j \in [\![1, r]\!] : d_j = k\}| = \delta_k - \delta_{k-1} =: \Delta_k$ où δ est défini dans le lemme 4.1.1.

Démonstration. On note $M = \mathcal{M}_{\mathcal{B}}(f)$, on a alors $\mathrm{rg}(M^k) = \sum_{j=1}^{r}(d_j - k)_+$. Donc $\Delta_k = \mathrm{rg}(M^{k+1}) + \mathrm{rg}(M^{k-1}) - 2\mathrm{rg}(M^k)$ d'après le théorème du rang.

D'où $\Delta_k = \sum_{j=1}^{r}(d_j - k - 1)_+ + (d_j - k + 1)_+ - 2(d_j - k)_+ = |\{j \in [\![1, r]\!] : d_j = k\}|$. □

La propriété montre que le nombre de d_j égaux à un entier donné se calcule en termes de f et des noyaux itérés.

Théorème 4.3.4 (Existence de la réduction de Jordan dans le cas nilpotent)**.** Pour tout $f \in \mathcal{L}(E)$ nilpotent il existe une base $\mathcal{B}$ telle que $\mathcal{M}_{\mathcal{B}}(f)$ est la matrice diagonale par blocs, les blocs étant $J_{d_1}, \ldots, J_{d_r}$ avec $\deg(\mu_f) = d_1 \geq \cdots \geq d_r$.

Démonstration. Par récurrence sur $n = \dim(E)$, la propriété est claire pour $n = 1$ et pour $n \geq 2$ tel que la propriété est vraie au rang n, on pose $x \in E$ tel que $\mu_{f,x} = \mu_f$ et soit F un supplémentaire stable de $E_{f,x}$, par hypothèse de récurrence sur $f|_F$, il existe $d_2 \geq \cdots \geq d_r$ et une base qui met $f|_F$ sous une forme de Jordan. Dans $E_{f,x}$, on prend la base $(x, fx, \ldots, f^{p-1}x)$ et dans cette base, $f|_{E_{f,x}}$ se met sous la forme d'un bloc de Jordan. De plus, comme $\mu_{f|_F}|\mu_f$, on a $\deg(\mu_f) = d_1 \geq d_2$. □

Remarque 4.3.5. Du cas nilpotent, on en déduit le cas général avec χ_f scindé par le lemme des noyaux.

Applications :

- Toute matrice M (avec χ_M scindé) est semblable à sa transposée (par une matrice symétrique).
- Pour tout $f \in \mathcal{L}(E)$, il existe φ une forme bilinéaire symétrique non dégénérée tel que f est autoadjoint pour φ.
- Le bicommutant de f est $k[f]$ ie $\{g \in \mathcal{L}(E) : \forall h \in \mathcal{L}(E), hf = fh \Rightarrow hg = gh\} = k[f]$.

Chapitre 5

20/10/2023 - Théorème d'Ascoli et applications – Arnaud Debussche

5.1 Le cas métrique

Soient (X, d_1) et (E, d_2) deux espaces métriques tels que X est compact. On s'intéresse à $\mathcal{C}(X, E)$. L'espace $\mathcal{C}(X, E)$ est métrique pour la métrique $d_\infty : f, g \in \mathcal{C}(X, E) \mapsto \sup_{x\in X} d_2(f(x), g(x))$.

Remarque 5.1.1. La convergence dans $(\mathcal{C}(X, E), d_\infty)$ implique la convergence simple des fonctions et $\mathcal{C}(X, E)$ est complet si et seulement si E est complet.

Question : Quels sont les compacts [1] de $\mathcal{C}(X, E)$?

Définition 5.1.2. Soit $H \subset \mathcal{C}(X, E)$.
- Pour $x \in X$, H est **équicontinue en** x si et seulement si pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $y \in X$, si $d_1(x, y) \leq \eta$, alors pour tout $f \in H$, $d_2(f(x), f(y)) \leq \varepsilon$.
- H est **équicontinue sur** X si et seulement si elle est équicontinue en tout $x \in X$.
- H est **uniformément équicontinue sur** X si et seulement si pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $x, y \in X$, si $d_1(x, y) \leq \eta$, alors pour tout $f \in H$, $d_2(f(x), f(y)) \leq \varepsilon$.

Remarque 5.1.3. Les deux premières définitions s'étendent dans un cadre non métrique mais pas la troisième.

Proposition 5.1.4 (Conséquence du théorème de Heine)**.** Soit X espace métrique compact, H est équicontinue sur X si et seulement si X est uniformément équicontinue sur X.

Proposition 5.1.5. Soit $f \in (\mathcal{C}(X, E))^{\mathbb{N}}$.
- Si $f_n \xrightarrow{d_\infty} f_\infty$, alors f est équicontinue.
- Si pour tout $x \in X$, $f_n(x) \xrightarrow{d_\infty} f_\infty(x)$ et f est équicontinue, alors $f \in \mathcal{C}(X, E)$ et $f_n \xrightarrow{d_\infty} f_\infty$.

Démonstration. Pour $\epsilon > 0$, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$, $\sup_{x\in X} d(f_n(x), f(x)) \leq \varepsilon$.

Pour tous $x \in X$ et $\varepsilon > 0$, il existe $\eta > 0$ tel que si $d(x, y) \leq \eta$ alors $d(f(x), f(y)) \leq \varepsilon$.

Pour $n \in [\![0, N-1]\!]$, il existe $\eta_n > 0$ tel que $d(x, y) < \eta_n$ implique $d(f_n(x), f_n(y)) \leq \varepsilon$.

Si on remplace η par $\min(\eta, \eta_0, \ldots, \eta_{N-1})$ alors si $d(x, y) \leq \eta$ implique pour tout $n \in \mathbb{N}$, $d(f_n(x), f_n(y)) \leq \varepsilon$.

Si f est équicontinue, pour tous $x \in X$, $n \in \mathbb{N}$ et $\varepsilon > 0$, il existe $\eta_x > 0$ tel que pour tout $x \in X$, si $d(x, y) \leq \eta$, alors $d(f_n(x), f_n(y)) \leq \varepsilon$. En faisant tendre n vers l'infini, on obtient $d(f(x), f(y)) \leq \varepsilon$.

Par compacité, il existe $x_1, \ldots, x_n \in X$ tels que $X \subset \bigcup_{i=1}^N B(x_i, \eta_{x_i})$. Alors pour $x \in X$, il existe $i \in [\![1, N]\!]$ tel que $x \in B(x_i, \eta_{x_i})$, donc $d(f_n(x), f(x)) \leq d(f_n(x), f_n(x_i)) + d(f_n(x_i), f(x_i)) + d(f(x_i), f(x)) \leq \varepsilon$ pour n assez grand. □

Théorème 5.1.6 (Théorème d'Ascoli (métrique))**.** Soit H partie de $\mathcal{C}(X, E)$ où (X, d) est un espace métrique compact et (E, d) est un espace métrique. Il y équivalence entre entre :

$$H \text{ est équicontinue sur } X \quad \text{et} \quad \forall x \in X, H_x = \{f(x), f \in H\} \text{est relativement compact dans } E$$

et

$$H \text{ est relativement compact dans } \mathcal{C}(X, E).$$

1. Les compacts sont utiles pour résoudre des équations

Remarque 5.1.7. Dans le cadre uniformément lipschitzien, on obtient facilement les conditions.

Démonstration. Le sens réciproque est le plus simple. Comme H est relativement compact, il est donc précompact ie pour tout $\varepsilon > 0$, il existe $f_1, \dots, f_N \in H$ tel que $H \subset \bigcup_{i=1}^N B(f_i, \epsilon/3)$. Pour tous $i \in [\![1, N]\!]$, $x \in X$ et $\varepsilon > 0$, il existe $\eta_i > 0$ tel que si $d(x, y) \leq \eta_i$, $d(f_i(x), f_i(y)) \leq \varepsilon/3$. En prenant η le minimum des η_i, on a alors si $d(x, y) \leq \eta$, pour tout $f \in H$, $d(f(x), f(y)) \leq \varepsilon$. Cela donne l'équicontinuité de H. Pour tout $x \in X$, avec l'application $\Phi_x : f \mapsto f(x)$, $H_x = \Phi_x(H)$ donc $\overline{H} \subset \Phi_x(\overline{H})$ donc est compact.

Pour le sens direct, soit $D \subset X$ un ensemble dénombrable, pour tout $x \in D$, il existe φ_x extractrice tel que $f_{\varphi_x(n)} \to y_x$. On fait une extraction diagonale : il existe donc φ extractrice indépendente de x tel que pour $x \in D$, $f_{\varphi(n)}(x) \to y_x$. On pose alors $y_x = f(x)$. H est uniformément équicontinue : pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que si $d(x, y) \leq \eta$, alors pour tout $f \in H$, $d(f(x), f(y)) \leq \varepsilon$. Soit $x \in X \setminus D$, il existe $\tilde{x} \in D$ tel que $d(x, \tilde{x}) \leq \eta$, donc $d(f_{\varphi(n+l)}(x), f_{\varphi(n)}(x)) \leq 3\varepsilon$ pour $n \geq N(\tilde{x})$, donc $(f_{\varphi(n)}(x))_n$ est une suite de Cauchy dans $\overline{H}_x$ compact donc $(f_{\varphi(n)}(x))_n$ est compacte donc elle converge simplement vers une fonction f et l'équicontinuité donne que $f \in \mathcal{C}(X, E)$ et $f_n \xrightarrow{d_\infty} f$. □

Exemples 5.1.8.
— L'ensemble $\{x \in [0, 1] \mapsto x^n, n \in \mathbb{N}\}$ n'est pas équicontinue en 1.
— Avec l'ensemble $\{f_n : x \in \mathbb{R}_+ \mapsto \sin(\sqrt{t + 4n^2\pi^2}), n \in \mathbb{N}\}$, pour $t \geq 0$, $f_n(t) = \sin(\sqrt{t + 4n^2\pi^2} - 2n\pi) = \sin\left(\frac{t}{\sqrt{t+4n^2\pi^2}+2n\pi}\right) \xrightarrow[n\to+\infty]{} 0$ donc les f_n convergent simplement vers 0 mais pas uniformément. Pour tous $n \in \mathbb{N}$ et $t \geq 0$, $|f_n'(t)| \leq \frac{1}{4n\pi}$ par le théorème des accroissments finis, f est équicontinue.
— Pour $X \subset \mathbb{R}^n$ et $E \subset \mathbb{R}^d$, pour $\alpha \in]0, 1[$, on pose

$$\mathcal{C}^\alpha(X, E) = \left\{ f : X \to E : \sup_{x,y \in X} \frac{\|f(x) - f(y)\|}{\|x - y\|^\alpha} < +\infty \right\}.$$

C'est un espace normé pour la norme $\|f\|_{\mathcal{C}^\alpha(X,E)} = \|f\|_{\mathcal{C}(X,E)} + |f|_{\mathcal{C}^\alpha(X,E)}$ où $|f|_{\mathcal{C}^\alpha(X,E)} = \sup_{x,y \in X} \frac{\|f(x)-f(y)\|}{\|x-y\|^\alpha}$. Si $\alpha = 1$, on le note plutôt $\mathrm{Lip}(X, E)$ ou $\mathcal{C}^{0,1}(X, E)$ (si $\alpha > 1$, il n'y aurait que des constantes). Les bornés B de $\mathcal{C}^\alpha(X, E)$ pour $\alpha \in]0, 1]$ sont équicontinues. En effet, pour M tel que $B \subset B_{\mathcal{C}^\alpha(X,E)}(0, M)$, si $f \in H$, $|f|_\alpha \leq M$, alors pour $\varepsilon > 0$, on peut prendre $\eta = \left(\frac{\varepsilon}{M}\right)^{\frac{1}{\alpha}}$. De plus, pour $x \in X$, $\|f(x)\| \leq \|f\|_{C(X,E)} \leq M$ donc $\{f(x), f \in B\}$ est relativement compact dans E donc d'après Ascoli les bornés de $\mathcal{C}^\alpha(X, E)$ sont compacts dans $\mathcal{C}(X, E)$ (on ne viole pas le théorème de Riesz puisque la topologie a été changée).

Remarque 5.1.9. Pour $0 < \alpha < \beta < 1$, il existe $M_{\alpha,\beta} > 0$ tel que pour tout $f \in \mathcal{C}^\beta(X, E)$, $\|f\|_{\mathcal{C}^\alpha(X,E)} \leq M_{\alpha,\beta} \|f\|^{\frac{\alpha}{\beta}}_{\mathcal{C}^\beta(X,E)} \|f\|^{1-\frac{\alpha}{\beta}}_{\mathcal{C}(X,E)}$.

Pour B borné de $\mathcal{C}^\beta(X, E)$ est compact dans $\mathcal{C}(X, E)$ et dans $\mathcal{C}^\alpha(X, E)$ pour $\alpha \in]0, \beta[$.

Soit $k \in \mathcal{C}([a, b] \times [a, b] \to \mathbb{R})$, on pose

$$K : \mathcal{C}([a, b]) \ni f \mapsto \left(Kf : x \mapsto \int_a^b k(x, y) f(y) \mathrm{d}y \right).$$

Si $K : \mathcal{C}([a, b]) \to \mathcal{C}([a, b])$, alors pour $f \in \overline{B}_{\mathcal{C}([a,b])}(0, 1)$, $|Kf(x_1) - Kf(x_2)| \leq \int_a^b |k(x_1, f) - k(x_2, f)||f(y)|\mathrm{d}y \leq \sup_{y \in [a,b]} |k(x_1, y) - k(x_2, y)||b - a| \leq \varepsilon$ pour $|x_1 - x_2| \leq \eta$ car k est uniformément continue sur $[a, b]$. De plus, $|Kf(x)| \leq \|k\|_\infty$ donc $K\overline{B}_{\mathcal{C}([a,b])}(0, 1)$ est relativement compact donc K est compact. Si $K : L^2([0, 1]) \to L^2([0, 1])$ alors pour tout $f \in L^2([a, b])$, $|Kf(x_1) - Kf(x_2)| \leq \sup_{y \in [a,b]} |k(x_1, y) - k(x_2, y)| \sqrt{b - a} \|f\|_2$ et $\|f(x)\| \leq \|K\|_\infty \sqrt{b - a} \|f\|_{L^2}$. Ainsi $K(\overline{K}_{L^2}(0, 1))$ vérifie le théorème d'Ascoli donc est relativement compact dans $\mathcal{C}^0([a, b])$ donc dans $L^2([a, b])$.

Remarque 5.1.10. Les opérateurs compacts servent pour le théorème de Schauder et le théorème de diagonalisation dans un Hilbert.

Remarque 5.1.11. Pour le théorème d'Ascoli,
— si E est inclus dans un espace vectoriel de dimenson finie, H_x est relativent compact si et seulement s'il est borné ;
— si on a l'équicontinuité, pour tout $x \in X$, H_x est relativement compact ssi $\{f(x), x \in H, f \in H\}$.
— Si X est connexe et équicontinue et E est inclus dans un espace de dimension finie, pour $x \in X$, H_x est relativement compact si et seulement si il existe $x_0 \in X$ tel que H_{x_0} relativement compact (X est non-nécessairement compact). Pour le montrer, on pose $A = \{x \in X : H_x \text{ borné}\}$. Pour le sens retour, $x_0 \in A$ donc $A \neq \emptyset$. A est ouvert car pour $x \in X$, il existe $\eta > 0$ tel que pour $f \in H$ si $d(x, y) \leq \eta$, alors $\|f(x) - f(y)\| \leq 1$ donc $\|f\|$ est bornée d'où $B(x, y) \subset A$. A est fermé car pour $x \in A^{\mathbb{N}}$ qui converge vers x_∞ dans X, il existe alors N tel que si $n \geq N$, $\|x_n - x\| \leq \eta$, alors pour tout $f \in H$, $\|f(x)\| \leq \|f(x_n)\| + 1$ donc $x \in A$. A est un ouvert fermé de X connexe donc $X = A$.

Théorème 5.1.12. Soit F un espace vectoriel normé tel que $F \subset \mathcal{C}^1([0,1])$ et F est borné pour la norme de $\mathcal{C}^0([0,1])$. Alors F est de dimension finie.

Démonstration. On rappelle que $\|f\|_{\mathcal{C}^0} = \sup_{[0,1]} |f|$ et $\|f\|_{\mathcal{C}^1} = \sup_{[0,1]} |f| + \sup_{[0,1]} |f'|$.

F est complet dans $\mathcal{C}^0([0,1])$ et dans $\mathcal{C}^1([0,1])$ car pour f suite de Cauchy pour $\|\cdot\|_{\mathcal{C}^1}$, alors f et f' est de Cauchy pour $\|\cdot\|_{\mathcal{C}^0}$. Il existe donc $f_\infty, g \in \mathcal{C}^0([0,1])$ tel que $f_n \to f_\infty$ et $f_n' \to g$ dans $\mathcal{C}^0([0,1])$. Nécessairement $f_\infty' = g$ donc $f_n \to f \in \mathcal{C}^1$.

L'application $i : (F, \|\cdot\|_{\mathcal{C}^1}) \ni f \mapsto f \in (F, \|\cdot\|_{\mathcal{C}^0})$ est bijective et continue (car 1-lipschitzien). D'après le théorème d'isomorphisme de Banach, il existe $M > 0$ tel que il est $f \in F$, $\|f\|_{\mathcal{C}^1} \leq M\|f\|_{\mathcal{C}^0}$. Soit B la boule de $(F, \mathcal{C}^0)$, soit $B \subset \overline{B}_{\mathcal{C}^1}(0, M)$, d'après le théorème d'Ascoli, $\overline{B}_{\mathcal{C}^1}(0, M)$ est relativement compact dans $\mathcal{C}^0$ donc B est relativement compact dans $\mathcal{C}^0([0,1])$. Finalement, d'après le théorème de Riesz, F est de dimension finie. □

Chapitre 6

MOPS Probabilités et statistiques 24/10/2023 - Estimation statistique – Jean-Christophe Breton

À partir
[Compléter avec notes du 17/10/2023]

Maximum de vraisemblance

On considère un modèle $(\mathcal{H}^n, (Q_\theta^{\otimes n})_{\theta\in\Theta})$ pour n observations iid. On suppose que le modèle est dominé par $\mu^{\otimes n}$.

Rappel. Pour un modèle "continu", en général $\mu = \lambda$ et alors $\frac{\mathrm{d}Q_\theta}{\mathrm{d}\mu} = f_\theta$ est la densité de probabilité usuelle et pour un modèle "discret" $\mu = \eta$ (où η est la mesure de comptage) et alors

$$Q_\theta(A) = \sum_{x\in A} q_\theta(x) = \int_A q_\theta(x)\eta(\mathrm{d}x)$$

où $q_\theta = \frac{\mathrm{d}Q_\theta}{\mathrm{d}\eta}$.

Exemple 6.0.1. Si $X \sim \mathcal{B}(p) \ll \delta_0 + \delta_1$ alors comme $\mathbb{P}(X = x)$ vaut p si $x = 1$ et vaut $1 - p$ si $x = 0$ ie

$$\mathbb{P}(X = x) = p^x(1 - p)^{1-x} = \frac{\mathrm{d}\mathcal{B}(p)}{\mathrm{d}(\delta_0 + \delta_1)}(x).$$

Définition 6.0.2. La **vraisemblance du modèle** $(\mathcal{H}^n, (Q_\theta^{\otimes n})_{\theta\in\Theta})$ est

$$L_n : \mathcal{H}^n \times \Theta \ni (x, \theta) \mapsto \frac{\mathrm{d}Q_\theta^{\otimes n}}{\mathrm{d}\mu^{\otimes n}}(x).$$

Comme on est iid, on a

$$L_n(x, \theta) = \prod_{i=1}^n L_1(x_i, \theta).$$

Exemple 6.0.3. Sur le modèle $\left(\{0, 1\}^n, (\mathcal{B}(p)^{\otimes n})_{p\in[0,1]}\right)$ est dominé par $\delta_0 + \delta_1$ et alors $L_n(x, \theta) = p^{\sum_{i=1}^n x_i}(1 - p)^{n-\sum_{i=1}^n x_i}$.

Exemple 6.0.4. Sur le modèle $(\mathbb{R}^n, (\mathcal{N}(m, \sigma^2)^{\otimes n})_{(m,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+})$ est dominé par λ et alors

$$L_n(x, \theta) = \frac{\exp\left(-\sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{n/2}}.$$

[cf. document de cours]
En général, on préfère manipuler la **log-vraisemblance** $l_n(x, \theta) = \ln L_n(x, \theta)$ ce qui donne dans notre cas $\sum_{i=1}^n l_1(x_i, \theta)$.

Définition 6.0.5. Pour un modèle statistique $(\mathcal{H}^n, (Q_\theta^{\otimes n})_{\theta\in\Theta})$ de vraisemblance $L_n(x,\cdot)$. On appelle **estimateur du maximum de vraisemblance** (EMV) toute statistique g_n à valeurs dans Θ telle que pour tout $x \in \mathcal{H}^n$,

$$L_n(x, g_n(x)) = \max_{\theta\in\Theta} L_n(x,\theta).$$

Proposition 6.0.6. Lorsque $L_n(x,\cdot)$ est différentiable et strictement positive, l'estimateur du maximum de vraisemblance $\widehat{\theta_n}$ est solution de $\nabla_\theta l_n(x,\theta) = 0$.

Démonstration. Si $\theta \mapsto L(x,\theta)$ est différentiable, son maximum est nécessairement un point critique, ça l'est aussi pour la log-vraisemblance. □

Exemple 6.0.7 (Modèle de Bernoulli)**.** On a $L_n(x,p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$ donc

$$l_n(x,p) = \left(\sum_{i=1}^n x_i\right)\ln p + \left(n - \sum_{i=1}^n x_i\right)\ln(1-p).$$

Alors $\partial_p l_n(x,p) = \dfrac{\sum_{i=1}^n x_i}{p} - \dfrac{n-\sum_{i=1}^n x_i}{1-p}$ qui a pour zéro $\widehat{p_n} = \overline{x_n}$. C'est bien un maximum qu'on a identifié puisque $\partial^2_{p,p} l_n(x,p) < 0$.

Exemple 6.0.8 (Modèle uniforme)**.** Sur le modèle $(\mathbb{R}_+^n, (\mathcal{U}([0,\theta])^{\otimes n})_{\theta>0})$ domin" par $\lambda|_{\mathbb{R}_+}$, on a $L_n(x,p) = \frac{1}{\theta^n}\mathbb{1}_{[0,\theta]}(\max x)$ donc $l_n(x,p) = -n\ln\theta + \ln(\mathbb{1}_{[0,\theta]}(\max x))$. Un dessin permet de voir simplement que le maximum de vraisemblance est réalisé en $\max(x)$ est donc l'EMV est $\widehat{\theta_n} = \max(X_1,\dots,X_n)$.

[Faire dessin]

Quelques qualités supplémentaires pour des estimateurs

Les notions d'efficacité et de bornes de Cramer-Rao, d'information de Fisher est à voir dans des références de statistiques. La notion suivante d'exhaustivité est déjà plus intéressante à étudier.

Définition 6.0.9. Une statistique g_n est dite **exhaustive** si et seulement si pour tout $\theta \in \Theta$, $\mathcal{L}_\theta(X_1,\dots,X_n | g_n(X_1,\dots,X_n))$ ne dépend pas de θ. Autrement dit connaître X ou connaître $g_n(X)$ apporte la même information.

Exemple 6.0.10 (Modèle de Bernoulli)**.** Dans le modèle de Bernoulli, $n\overline{X_n}$ est une statistique exhaustive. En effet, pour $p \in [0,1]$, on étudie $\mathcal{L}_p(X|\overline{X_n})$. $\mathcal{B}(p)^{\otimes n}(X_1 = y_1,\dots,X_n = y_n | n\overline{X_n} = z)$ vaut 0 si $y_1+\cdots+y_n \neq z$. Sinon cela vaut

$$\frac{\mathcal{B}(p)^{\otimes n}(X_1=y_1,\dots,X_n=y_n, n\overline{X_n}=z)}{\mathcal{B}(p)^{\otimes n}(n\overline{X_n}=z)} = \frac{\prod_{i=1}^n \mathcal{B}(p)(X_i=y_i)}{\mathcal{B}(p)^{\otimes n}(n\overline{X_n}=z)} = \frac{\prod_{i=1}^n p^{x_i}(1-p)^{1-y_i}}{\binom{n}{z}p^z(1-p)^{n-z}} = \frac{1}{\binom{n}{z}}.$$

Théorème 6.0.11 (de Neymann-Fisher)**.** Une statistique g_n est exhaustive si et seulement si la vraisemblance s'écrit sous la forme $L_n(x,\theta) = \psi(g_n(x),\theta)\gamma(x)$.

Une autre notion intéressante est celle de la robustesse. Aucun définition formelle ne sera donnée ici. Un procédure statistique est qualifiée de robuste quand ses performances (les propriétés des estimateurs) sont peu sensibles à de petites déviations des hypothèses de base sur le modèle.

Exemple 6.0.12. Avec un échantillon $(1,2;\ 1,3;\ 0,8;\ 1,15;\ 1,13;\ 0,95;\ 121;\ 1,31)$, la moyenne prouve ne pas être robuste alors que la médiane l'est plus.

Pour estimer un paramètre inconnu, l'estimateur statistique peut être complété d'un intervalle de confiance.

Exemple 6.0.13. Pour le modèle $(\mathbb{R}^n, (\mathcal{N}(m,1)^{\otimes n})_{m\in\mathbb{R}})$, un bon estimateur est $\overline{X_n}$ lorsque $X \sim \mathcal{N}(m,1)^{\otimes n}$. On sait que $\overline{X_n} \sim \mathcal{N}\left(m, \frac{1}{n}\right)$ ie $\sqrt{n}(\overline{X_n}-m) \sim \mathcal{N}(0,1)$. Ainsi $\mathbb{P}_m\left(|\sqrt{n}(\overline{X_n}-m)| > q\right) = \mathbb{P}(|N| > q_{1-\frac{\alpha}{2}}) = \alpha$ où $N \sim \mathcal{N}(0,1)$ ainsi $\mathbb{P}_m\left(m \in \left[\overline{X_n} - \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \overline{X_n} + \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right]\right) = 1-\alpha$. On pose alors $IC_{1-\alpha}(m) = \left[\overline{X_n} - \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \overline{X_n} + \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right]$ est un intervalle de confiance pour m de niveau $1-\alpha$ avec α en général valant 5%, 1% ou 1‰.

Définition 6.0.14. Soit $\alpha \in]0,1[$ donné, on appelle intervalle de confiance ou région de confiance de niveau $1-\alpha$ la famille non vide de parties $C_{x_1,\dots,x_n}$ de $\Theta \subset \mathbb{R}^d$ tel que pour tout $\theta \in \Theta$, $\mathbb{P}_\theta(\theta \in C_X) \leq 1-\alpha$.

Intervalle de confiance gaussien

[Rappel loi du χ^2 et loi de Student, cf. Moodle]

Théorème 6.0.15 (de Fisher)**.** Pour $X \sim \mathcal{N}(m, \sigma^2)^{\otimes n}$ où $\sigma > 0$, on pose $\overline{X_n}$ (resp. S_n^2) la moyenne empirique (resp. la variance empirique). Ainsi $\overline{X_n} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, $\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$, $\overline{X_n} \perp\!\!\!\perp \frac{nS_n^2}{\sigma^2}$ et $\frac{\overline{X_n}-m}{\sqrt{S_n^2}}\sqrt{n-1} \sim \mathcal{T}(n-1)$.

Intervalle de confiance pour la moyenne

Si σ^2 est connue, on a alors $\sqrt{n}\frac{\overline{X_n}-m}{\sigma} \sim \mathcal{N}(0,1)$, on a alors comme intervalle de confiance pour m de niveau de risque $1-\alpha$ $IC_{1-\alpha}(m) = \left[\overline{X_n} - q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X_n} + q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ (où $q.$ est le quantile pour la loi normale).

Si σ^2 est inconnue, on sait d'après le théorème de Fisher que $\frac{\overline{X_n}-m}{\sqrt{S_n^2}}\sqrt{n-1} \sim \mathcal{T}(n-1)$ donc $\mathbb{P}_m(\sqrt{n-1}\left|\frac{\overline{X_n}-m}{\sqrt{S_n^2}}\right| \leq q_{1-\alpha/2}) = 1-\alpha$ (où $q_{1-\alpha/2}$ est le quantile pour la loi de Student). Ainsi un intervalle de confiance pour m de niveau de risque $1-\alpha$ est $IC_{1-\alpha}(m) = \left[\overline{X_n} - q_{1-\alpha/2}\frac{\sqrt{S_n^2}}{\sqrt{n-1}}, \overline{X_n} + q_{1-\alpha/2}\frac{\sqrt{S_n^2}}{\sqrt{n-1}}\right]$.

Intervalle de confiance pour la variance

Si la moyenne m est connue, on pose $\frac{1}{n}T = \sum_{i=1}^n (X_i - m)^2$ alors $n\frac{T}{\sigma^2} \sim \chi^2(n)$. Comme on peut le voir sur un dessin, on a alors comme intervalle de confiance pour σ avec niveau $1-\alpha$ $IC_{1-\alpha}(\sigma) = \left[\sqrt{\frac{nT}{q_{1-\alpha/2}}}, \sqrt{\frac{nT}{q_{\alpha/2}}}\right]$ (où les $q.$ sont des quantiles pour la loi du χ^2 pour n degrés de libertés).

Si la moyenne m est inconnue, alors $\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$, ainsi un intervalle de confiance pour σ de niveau $1-\alpha$ est $IC_{1-\alpha}(\sigma) = \left[\sqrt{\frac{nS_n^2}{q_{1-\alpha/2}}}, \sqrt{\frac{nS_n^2}{q_{\alpha/2}}}\right]$ (où les $q.$ sont des quantiles pour la loi du χ^2 pour $n-1$ degrés de libertés).

Intervalle de confiance pour la moyenne de grands échantillons

On aimerait se ramener au cas gaussiens grâce au TLC.

Lorsque la variance est connue, $\mathbb{P}\left(\sqrt{n}\left|\frac{\overline{X_n}-m}{\sigma}\right| \leq q_{1-\alpha/2}\right) \xrightarrow[n\to+\infty]{} \mathbb{P}(|N| \leq q_{1-\alpha/2}) = 1-\alpha$. Pour n "grand" (typiquement $n \geq 30$), un intervalle de confiance asymptotique de m de niveau $1-\alpha$ est $IC_{1-\alpha}(m) = \left[\overline{X_n} - q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X_n} + q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$.

Lorsque la variance est inconnue, le TLC donne toujours $\sqrt{n}\frac{X_n-m}{\sigma} \Rightarrow \mathcal{N}(0,1)$ et de plus $S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$ (car S_n^2 est consistant) donc $\frac{\sigma}{\sqrt{S_n^2}} \xrightarrow{\mathbb{P}} 1$. En utilisant le lemme de Slutsky et le continuous mapping theorem, $\sqrt{n}\frac{X_n-m}{\sqrt{S_n^2}} \Rightarrow \mathcal{N}(0,1)$. Pour n "grand", un intervalle de confiance asymptotique de m de niveau $1-\alpha$ est $IC_{1-\alpha}(m) = \left[\overline{X_n} - q_{1-\alpha/2}\frac{\sqrt{S_n^2}}{\sqrt{n}}, \overline{X_n} + q_{1-\alpha/2}\frac{\sqrt{S_n^2}}{\sqrt{n}}\right]$.

Exemple 6.0.16. Sur un modèle exponentiel $(\mathbb{R}_+^n, (\mathcal{E}(\theta)^{\otimes n})_{\theta>0})$, un estimateur du paramètre est $\widehat{\theta_n} = \frac{1}{\overline{X_n}}$. On a $\mathbb{E}_\theta[X] = \frac{1}{\theta}$ et $\mathbb{V}_\theta[X] = \frac{1}{\theta^2}$. On sait que $\sqrt{n}\frac{\overline{X_n}-\frac{1}{\theta}}{\sqrt{\frac{1}{\theta^2}}} = \sqrt{n}(\theta\overline{X_n} - 1) \Rightarrow \mathcal{N}(0,1)$. Ainsi pour n assez grand, un intervalle de confiance asymptotique de θ de niveau $1-\alpha$ est $IC_{1-\alpha}(m) = \left[\frac{1-\frac{q_{1-\alpha/2}}{\sqrt{n}}}{\overline{X_n}}, \frac{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}}{\overline{X_n}}\right]$.

Exemple 6.0.17. Pour le modèle de Bernoulli, un premier intervalle de confiance est donné par l'inégalité de Bienaymé-Tchebychev. Pour $t > 0$, elle donne $\mathbb{P}(|\overline{X_n} - p| > t) \leq \frac{1}{4nt^2}$ et un intervalle de confiance pour p de niveau $1-\alpha$ est $IC_{1-\alpha}(m) = \left[\overline{X_n} - \frac{1}{\sqrt{4n\alpha}}, \overline{X_n} + \frac{1}{\sqrt{4n\alpha}}\right]$.

Un deuxième intervalle de confiance pour m de niveau $1-\alpha$ est donné par asymptotiquement par [cf. document] [faire les calculs sans utiliser Slutsky (en utilisant l'expression exacte de la variance) et en utilisant Slutsky]

Un dernier intervalle de confiance pour m est donnée par l'inégalité de Hoeffding, car $\mathbb{P}([\overline{X_n} - m| > \epsilon) \leq 2\exp\left(-\frac{n\epsilon^2}{2}\right)$ donc un intervalle de confiance pour m de niveau de confiance $1-\alpha$ est

$$IC_{1-\alpha}(m) = \left[\overline{X_n} - \sqrt{\frac{2}{n}\ln\left(\frac{2}{\alpha}\right)}, \overline{X_n} + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\alpha}\right)}\right].$$

Chapitre 7

Cours Mercredi après-midi 25/10/2023 - Matrices sur les corps finis – Matthieu Romagny

Références : [Per96]

7.1 $GL_n(\mathbb{F}_q)$ et ses proches amis

Théorème 7.1.1. Soit q une puissance d'un nombre premier, on a alors pour tout $n \in \mathbb{N}^*$

$$\gamma_n := |GL_n(\mathbb{F}_q)| = (q^n - 1)\cdots(q^n - q^{n-1}) = q^{\frac{n(n-1)}{2}}(q^n - 1)\cdots(q-1),$$
$$|SL_n(\mathbb{F}_q)| = \frac{\gamma_n}{q-1} = |PGL_n(\mathbb{F}_q)| \quad \text{et} \quad |PSL_n(\mathbb{F}_q)|\frac{\gamma_n}{d(q-1)}$$

où $d = n \wedge (q-1)$.

Applications :

1. Isomorphismes exceptionnels ([Per96]) : $PSL_2(\mathbb{F}_2) \simeq \mathfrak{S}_3$ et $PSL_2(\mathbb{F}_3) = \mathfrak{A}_4$ qui sont non-simples. Tous les autres $PSL_n(k)$ sont simples.
2. $\underset{|G|=n}{G} \underset{\text{Cayley}}{\hookrightarrow} \mathfrak{S}_G \underset{\text{mat. de perm.}}{\overset{\forall p}{\hookrightarrow}} GL_n(\mathbb{F}_q) \hookleftarrow$ mat. diag.. On démontre ainsi l'existence d'un p-Sylow pour G.

Remarque 7.1.2. $\frac{\gamma_n}{\gamma_{n-1}} = [GL_n(\mathbb{F}_q) : GL_{n-1}(\mathbb{F}_q)] = q^{n-1}(q^n - 1)$.

7.2 Matrices nilpotentes

(décomposition de Fitting)

L'ensemble $Nil(E) \subset \mathcal{L}(E)$ des endomorphismes nilpotents n'est pas stable par addition, en revanche, $Nil(E)$ est stabkle par multiplication par $\lambda \in k$.

[Exercices dans Calders et Germoni]

Théorème 7.2.1. Le nombre d'endomorphismes nilpotents de $E = \mathbb{F}_q^n$ est $\nu_n = q^{n(n-1)}$

Démonstration. La décomposition de Fitting associe à tout $f \in \mathcal{L}(E)$ une unique décomposition $E = F \oplus G$ où F et G sont f-stables et tels que $f|_F \in Nil(F)$ et $f|_G \in GL(G)$. On obtient alors une bijection

$$L(E) \simeq \bigsqcup_{F \oplus G = E} Nil(F) \times GL(G).$$

Lemme 7.2.2. L'ensemble des décompositions $F = E_1 \oplus \cdots \oplus E_r$ en sous-espaces de dimensions fixés $n_i = \dim(E_i)$ est en bijection avec

$$\frac{GL_n(\mathbb{F}_q)}{GL_{n_1}(\mathbb{F}_q) \times \cdots \times GL_{n_r}(\mathbb{F}_q)}$$

(qui est un sous groupe de matrices diagonales par blocs) de cardinal $\dfrac{\gamma_n}{\gamma_{n_1}\cdots\gamma_{n_r}}$.

Démonstration. L'ensemble des ces décompositions admet une action de $G = GL_n(\mathbb{F}_q)$ par

$$g(E_1, \dots, E_r) = (gE_1, \dots, gE_r).$$

Cette action est transitive car ...
Le stabilisateur de la décomposition $E = \bigoplus_{i=1}^r E_i$ est le sous-groupe $\prod_{i=1}^n (GL_{n_1}(\mathbb{F}_q)) \hookrightarrow GL_n(\mathbb{F}_q)$ des matrices diagonales par blocs. La relation orbite/stabilisateur fournité la bijection de l'énoncé. □

Ainsi

$$\mathcal{L}(E) = \bigsqcup_{k=0}^{n} \bigsqcup_{F \oplus G = E : \dim(F) = k, \dim(G) = n-k} Nil(F) \times GL(G).$$

On prend les cardinaux $q^{n^2} = \sum_{k=0}^n \frac{\gamma_n}{\gamma_k \gamma_{n-k}} \nu_k \gamma_{n-k}$ et on en déduit que $\frac{q^{n^2}}{\gamma_n} = \sum_{k=0}^n \frac{\nu_k}{\gamma_k}$ ainsi $\frac{\nu_n}{\gamma_n} = \frac{q^{n^2}}{\gamma_n} - \frac{q^{(n-1)^2}}{\gamma_{n-1}}$ et ceci donne le résultat. □

7.3 Matrices diagonalisables et trigonalisables

(partition et actions de groupes)

Théorème 7.3.1. Le nombre de matrices diagonalisables (resp. trigonalisables) dans $GL_n(\mathbb{F}_q)$ est égal à

$$\sum_{n_1 + \dots + n_{q-1}} \frac{\gamma_n}{\gamma_{n_1} + \dots + \gamma_{n_{q_1}}} \quad \left(\text{resp.} \quad \sum_{n_1 + \dots + n_{q-1}} \frac{\gamma_n}{\gamma_{n_1} + \dots + \gamma_{n_{q_1}}} q^{\sum n_i^2} - n \right).$$

Démonstration. La décomposition en sous-espaces caractéristiques d'une matrice trigonalisables (dont les valeurs propres sont parmi les éléments $\zeta_1, \dots, \zeta_{q-1} \in \mathbb{F}_q^\times$) associe à f une décomposition $E = E_1 \oplus \dots \oplus E_{q-1}$ (où les E_i sont les sous espaces caractéristiques de ζ_i) et des endomorphismes nilpotents $f|_{E_i} - \zeta_i \mathrm{id}_{E_i} = g_i$.
On obtient ainsi une bijection de l'ensemble des matrices diagonalisables de $GL_n(\mathbb{F}_q)$ sur

$$\bigsqcup_{E_1 \oplus \dots \oplus E_{q-1} = E} Nil(E_1) \cdots \times \cdots Nil(E_{q-1})$$

de réciproque $(g_1, \dots, g_r) \mapsto f_g$ où $f_g|_{E_i} = g_i + \zeta_i \mathrm{id}_{E_i}$ et le calcul de cardinal nous donne le résultat. □

Remarque 7.3.2. La somme est indicée par les $(q-1)$-uplets d'entiers positifs de somme n. C'est différent de la notion de partition de n : une partition est non ordonnée et sans partie nulle (ie 1,2,2 et 2,1,2 sont les mêmes partitions de 5 et 0,5 et 5 sont les mêmes partitions de 5).

Question (ouverte) : combien y a-t-il d'endomorphismes semi-simples dans $GL_n(\mathbb{F}_q)$ c'est-à-dire de matrices qui deviennent diagonalisables sur une extension $\mathbb{F}_{q^N}$.

Remarque 7.3.3. La décomposition de Jordan-Chevalley donne : si $S(N)$ est l'ensemble des endomorphismes semi-simples qui commutent avec N, $M_n(\mathbb{F}_q)$ est en bijection avec $\bigsqcup_{N \in Nil} S(N)$ par M donne sa composante semi-simple S dans sa décomposition $M = S + N$.

7.4 Générateurs de GL_n : dilatations et transvections

(actions de groupes)

Théorème 7.4.1. Le nombre de transvections de $GL_n(\mathbb{F}_q)$ est

$$\frac{(q^n - 1)(q^{n-1} - 1)}{q - 1} = (q^{n-1})(1 + q + \dots + q^{n-1}).$$

Démonstration. Soit T l'ensemble des transvections de $E = (\mathbb{F}_q)^n$. $GL(E)$ agit par conjugaison sur T, transitivement car toutes les transvections ont pour matrice $\begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & 0 \\ 0 & I_{n-2} \end{pmatrix}$ dans une certaine base.
Il reste à calculer le stabilisateur (ici : le commutant) d'une transvection $u(x) = x + f(x)a$ où $f \in E^* \setminus \{0\}$, $a \in \ker(f) =: H$ et $a \neq 0$. Soit $g \in GL(E)$ qui commute avec u alors (après calcul) on voit que $g(H) \subset H$ d'où $g(H) = H$ par égalité des dimensions car $g \in GL(E)$.

Par la suite, on va fixer $b \notin H$n donc E aura une base $\mathcal{B} = \{a, a_2, \dots, a_{n-1}, b\}$ avec $a, a_2, \dots, a_{n-1} \in H$. En prenant $x = b$, on a $f(b)g(a) = f(g(b))a$ donc $g(a) = \lambda a$ pour $\lambda = \frac{f(g(b))}{f(b)} \in k$.

Posons $v = \lambda^{-1}g$, on obtient après calcul que $v(b) - b \in H$. Finalement $v(a) = a$, $v(H) = H$ et $v(b) - b \in H$.

Réciproquement, si v vérifie ces conditions et $\lambda \in k^*$ alors on vérifie que $g := \lambda v$ commute avec u. Ainsi

$$Mat_{\mathcal{B}}(v) = \left(\begin{array}{c|ccc|c} 1 & & * & & * \\ \hline 0 & & & & \\ \vdots & & * & & * \\ 0 & & & & \\ \hline 0 & 0 & \dots & 0 & 1 \end{array}\right).$$

Finalement, le cardinal de l'ensemble des commutants vaut $(q-1)\gamma_{n-2}q^{2(n-2)+1}$ et donc on obtient le résultat. □

Théorème 7.4.2. Le nombre de dilatations de $GL_n(\mathbb{F}_q)$ est égal à $(q-2)q^{n-1}(1 + \dots + q^{n-1})$.

Démonstration. Une dilatation u est déterminée par la donnée de $\lambda = \det(u)$ et la décomposition $E = D \oplus H$ où $H = \ker(u - \mathrm{id})$ et D est la droite propre pour λ donc le cardinal vaut nécessairement $(q-2)\frac{\gamma_n}{\gamma_1\gamma_{n-1}}$ (pour que le supplémentaire soit en somme directe on ne peut pas avoir $\lambda = 1$) ce qui donne le résultat. □

Chapitre 8

MOPS Probabilités et statistiques 26/10/2023 - Chaînes de Markov à espace d'états dénombrable – Nathalie Krell

8.1 Introduction

Soit E un ensemble dénombrable et soit X_0 un état initial.

Définition 8.1.1. Le processus $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène ssi pour tout $y, x_0, \ldots, x_n \in E$ pour $n \geq 0$, les probabilités suivantes ne dépendent que de x_n et y, $\mathbb{P}(X_{n+1} = y | X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = y | X_n = x_n) = \mathbb{P}(X_1 = y | X_0 = x_n)$. On note $P(x, y)$ (ou $P_{x,y}$) cette quantité que l'on appelera probabilité de transition de l'état x à y.

Proposition 8.1.2. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov, alors il existe $f : E \times F \to E$ mesurable et $(U_n)_{n\geq 0}$ iid de va à valeurs dans $(E, \mathcal{F})$, indépendant de X_0 tels que $X_{n+1} = f(X_n, U_{n+1})$.

Réciproquement pour X processus aléatoire s'il existe $f : E \times F \to E$ mesurable et $(U_n)_{n\geq 0}$ iid de va à valeurs dans $(E, \mathcal{F})$, indépendant de X_0 tels que $X_{n+1} = f(X_n, U_{n+1})$ alors X est une chaîne de Markov homogène à valeurs dans E.

Exemple 8.1.3. Avec $E = \mathbb{N}$, $F = \{0, 1\}$ et $f(x, u) = u(x + u)$.

Exemple 8.1.4 (Marche aléatoire simple)**.** ["$X_{n+1} = X_n + \mathcal{R}(p)$", écrire la matrice de transition] Si $p = 1/2$ la marche aléatoire est symétrique et asymétrique sinon.

Exemple 8.1.5 (Marche aléatoire sur $\mathbb{Z}^d$)**.** $P_{x,y} = \frac{\delta_{|x-y|,1}}{d}$.

8.2 Loi d'une chaîne de Markov

Théorème 8.2.1. Soit X une chaîne de Markov homogène de matrice de transition P et de loi initiale μ_0, alors $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \mu(x_0) P_{x_0,x_1} \cdots P_{x_{n-1},x_n}$.

On posera par la suite $\mu_n = \mu_0 P^n$.

Remarque 8.2.2. La loi de la chaîne de Markov X homogène est entièrement déterminé par μ_0 et P.

Théorème 8.2.3 (Propriété de Markov faible)**.** Soit X une chaîne de Markov homogène de matrice de transition P et de loi initiale μ_0, alors pour tout $k \geq 0$ conditionnellement à $\{X_k = x\}$, $(X_{k+n})_{n\geq 0}$ est une chaîne de Markov de matrice de transition P est de loi initiale $\delta_x \perp\!\!\!\perp X_0, \ldots, X_k$.

Théorème 8.2.4 (Propriété de Markov forte)**.** Soit T un temps d'arrêt adapté à la filtration naturelle de X conditionnellement à $T < +\infty$ et $X_T = x$, alors $(X_{T+n})_{n\geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale $\delta_x \perp\!\!\!\perp X_0, \ldots, X_T$.

8.3 Récurrence et transience

On pose $T_x^1 = T_x = \inf\{k \geq 1 : X_k = x\}$ et $T_x^{n+1} = \inf\{k > T_x^n : X_k = x\}$.

Définition 8.3.1. Soit X une chaîne de Markov partant de x.
L'état x est dit transient (resp. récurrent) ssi $\mathbb{P}_x(T_x < +\infty) < 1$ (resp. $\mathbb{P}_x(T_x < +\infty) = 1$).
Si x est récurrent, on dit que x est récurrent positif (resp. récurrent nul) ssi $\mathbb{E}_x(T_x) = +\infty$ (resp. $\mathbb{E}_x(T_x) < +\infty$).

8.4 Temps de retour et nombre de visite

On note $N_x^n = \sum_{k=0}^n \mathbb{1}_{X_k=x}$ et $N_x = \sum_{k\geq 0} \mathbb{1}_{X_k=x}$. Ainsi $N_x^n \geq p+1$ ssi $T_x^p \leq n$ et $N_x \geq p+1$ ssi $T_x^p < +\infty$.

Proposition 8.4.1. Soit $x \in E$ et soit X une chaîne de Markov partant de x alors :

1. si $T_x^n < +\infty$, les va $T_x, T_x^2 - T_x, \ldots, T_x^{n+1} - T_x^n$ sont iid,
2. $\lim_{n\to+\infty} \frac{1}{n} N_n^x = \frac{1}{\mathbb{E}_x(T_x)}$.

Démonstration. Soit $T \sim T_x^{\otimes\mathbb{N}}$, on va montrer que $(T_x, \ldots, T_x^n) \sim (I_1, \ldots, I_1 + \cdots + I_n)$.
On note $E(s,t) = \{x \notin \{X_s, \ldots, X_t\}\}$. On a donc $\{T_x = t_1, \ldots, T_x^n = t_n\} = E(s,t_1) \cup \cdots \cup E(t_{n-1}+, t_n)$ d'après la propriété de Markov, les ensembles à droites de l'égalité sont indépendents.
Par ailleurs, la chaîne est homogène donc $\mathbb{P}(E(t_{i-1}+1, t_i)) = \mathbb{P}(E(1, t_i - t_{i-1})) = \mathbb{P}(I_1 = t_i - t_{i-1})$ d'où $\mathbb{P}(T_x = t_1, \ldots, T_x^n = t_n) = \mathbb{P}(I_1 = t_1)\mathbb{P}(I_2 = t_2 - t_1) \cdots \mathbb{P}(I_n = t_n - t_{n-1})$ ce qui donne le résultat par iid. □

Corollaire 8.4.2.

1. Si x est récurrent, X revient preque sûrement une infinité de fois à son état initiale ie $\mathbb{P}_x(N_x = +\infty) = 1$.
2. Si x est transient, presque sûrement X visite x un nombre fini de fois. Le nombre de visites partant du point x suit la loi géométrique de paramètre $a = \mathbb{P}_x(T_x = +\infty)$ et pour $k \geq 1$, $\mathbb{P}_x(N_x = k) = \mathbb{P}_x(T_x = +\infty)(1 - \mathbb{P}_x(T_x = +\infty))^{k-1}$.

Démonstration. D'après la proposition précédente, $\mathbb{P}_x(N_x > k) = \mathbb{P}_x(T_x < +\infty, \ldots, T_x^k < +\infty) = \mathbb{P}_x(T_x < +\infty, T_x^2 - T_x, \ldots, T_x^k - T_x^{k-1} < +\infty) = \mathbb{P}_x(T_x < +\infty)^k$. □

Proposition 8.4.3. Si P est irréductible,

1. tous les états sont de la même nature (récurrents positifs, récurrents nuls ou transients).
2. dans le cas récurrent, tous les points de E sont visités infiniment souvent.
3. dans le cas transient, les sous-ensembles finis de E ne sont visités un nombre fini de fois.

Conséquence : Une chaîne de Markov finie et irréductible est récurrente positive.

Chapitre 9

Cours Mercredi après-midi 2 08/11/2023 - Formule du changement de variables – Jürgen Angst

[Voir les notes du journal de bord, notamment aire et co-aire]

Théorème 9.0.1 (Formule du changement de variables)**.** Soient U un ouvert de $\mathbb{R}^n$ et f un $\mathcal{C}^1$-difféomorphisme de U dans $f(U)$, alors g est intégrable sur $f(U)$ ssi $x \mapsto (g \circ f)(x) \cdot |\det D_x f|$ est intégrable sur U. Dans ce cas,

$$\int_{f(U)} g(y)\mathrm{d}y = \int_U g(f(x))|\det D_x f|\mathrm{d}x$$

où $D_x f$ est la différentielle de f en x.

Exemple 9.0.2. Le changement de variable polaire est le changement de variables suivant : $U = \mathbb{R}_+^* \times]0, 2\pi[$, $f(U) = \mathbb{R}^2 \setminus \{0\}$ et $f(r, \theta) = (r\cos\theta, r\sin\theta)$. Alors

$$D_{(r,\theta)} f = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix}$$

d'où $\left|\det D_{(r,\theta)}\right| = r$ ie si g est L^1 sur $\mathbb{R}^2$ alors

$$\int_{\mathbb{R}^2} g(x, y)\mathrm{d}x\mathrm{d}y = \int_0^{+\infty} \int_0^{2\pi} g(r\cos\theta, r\sin\theta) r \mathrm{d}r\mathrm{d}\theta.$$

Si on cherche à calculer $I = \displaystyle\int_{\mathbb{R}} \exp\left(-\frac{x^2}{2}\right)\mathrm{d}x$, on a

$$I^2 = \int_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right)\mathrm{d}x\mathrm{d}y = \int_0^{+\infty}\int_0^{2\pi} \exp\left(-\frac{r^2}{2}\right) r\,\mathrm{d}r\,\mathrm{d}\theta = 2\pi,$$

ainsi $I = \sqrt{2\pi}$.

Preuve du théorème 9.0.1. Comme les deux membres de l'égalité sont linéaires en g, on peut se ramener au cas où g est étagée et par suite au cas où g est l'indicatrice de l'image d'un compact K de U ie $g = \mathbb{1}_{f(K)}$. On veut donc montrer que

$$\lambda_n(f(K)) = \int_K |\det D_x f|\lambda_n(\mathrm{d}x).$$

Dans le cas où f est linéaire inversible *ie* $f(x) = Ax$, on a $D_x f = A$ et $|\det D_x f| = |\det A|$.
La décomposition polaire de A est $S = \left(A^T A\right)^{1/2} \in \mathcal{S}_n^+(\mathbb{R})$ et $A = VS$ où $V = AS^{-1} \in \mathcal{O}_n(\mathbb{R})$.
On diagonalise $S = QDQ^T$ avec $Q \in \mathcal{O}_n(\mathbb{R})$ et $D = \mathrm{diag}(\mu_1, \ldots, \mu_n)$ où les μ_i sont positifs, de sorte que $A = VQDQ^T$ et $VQ, Q^T \in \mathcal{O}_n(\mathbb{R})$.
Comme la mesure de Lebesgue est invariante sous l'action de $\mathcal{O}_n(\mathbb{R})$, on a

$$\lambda_n(AK) = \lambda_n\left(VQDQ^T K\right) = \lambda_n\left(DQ^T K\right) = \mu_1 \cdots \mu_n \cdot \lambda_n\left(Q^T K\right) = (\det D)\lambda_n(K) = |\det A|\lambda_n(K).$$

Dans le cas général, il s'agit de linéariser f. On se donne $\varepsilon, \delta > 0$ et on recouvre K par des cubes disjoints B_i de côtés de longueurs ε et de centre c_i, on note alors $K_\varepsilon = \bigcup_i B_i$ le ε-recouvrement correspondant.
D'après la formule de Taylor, pour $c_i + h \in B_i$, $f(c_i + h) = f(c_i) + D_{c_i} f(h) + o(\|h\|)$.
Comme $x \mapsto D_x f$ est uniformément continue sur K, uniformément en i, on a alors, quitte à réduire ε, $f(B_i) \subset f(c_i) + (1+\delta) D_{c_i} f(B_i)$.
Ainsi,

$$\begin{aligned}\lambda_n\left(f\left(B_i\right)\right) \leq \lambda_n\left(f\left(c_i\right) + (1+\delta) \cdot D_{c_i} f\left(B_i\right)\right) = \lambda_n\left((1+\delta) \cdot D_{c_i} f\left(B_i\right)\right) = (1+\delta)^n \cdot \lambda_n\left(D_{c_i} f\left(B_i\right)\right) \\ = (1+\delta)^n \cdot |\det D_{c_i} f| \, \lambda_n(B_i).\end{aligned}$$

Toujours par uniforme continuité de $x \mapsto D_x f$, on a $|\det D_{c_i} f| \leq (1+\delta) |\det D_x f|$ pour tout $x \in B_i$ uniformément en i.
En intégrant sur B_i,

$$|\det D_{c_i} f| \lambda_n(B_i) \subset (1+\delta) \int_{B_i} |\det D_x f| \, \lambda_n(\mathrm{d}x).$$

Ainsi, en sommant sur i, on obtient

$$\lambda_n(f(K)) \leq \sum_i \lambda_n(f(B_i)) \leq (1+\delta)^{n+1} \sum_i \int_{B_i} |\det D_x f| \ \lambda_n(\mathrm{d}x) = (1+\delta)^{n+1} \int |\det D_x f| \, \mathbb{1}_{K_\varepsilon}(x) \, \lambda_n(\mathrm{d}x),$$

par théorème de convergence dominée lorque $\varepsilon \to 0$, on obtient $\lambda_n(f(K)) \leq (1+\delta)^{n+1} \int_K |\det D_x f| \lambda_n(\mathrm{d}x)$.
En faisant tendre δ vers 0, on en conclut que $\lambda_n(f(K)) \leq \int_K |\det D_x f| \lambda_n(\mathrm{d}x)$.
Pour l'inégalité inverse, on peut utiliser l'inégalité ci-dessus pour f^{-1}. □

[mesure de Haar]
[Application : $(X, Y) \sim \Gamma(a, \lambda) \otimes \Gamma(b, \lambda)$, alors $(X+Y, \frac{X}{X+Y}) \sim \Gamma(a+b, \lambda) \otimes \beta(a, b)$]

Chapitre 10

09/11/2023 - Gauss – Jean-Christophe Breton

[Rattraper le cours du 09/11/2023]

Proposition 10.0.1.

Démonstration. □

Contre-exemple 10.0.2.

Théorème 10.0.3 (TLC multidimensionel)**.** Soit $(X^{(n)})_{n\geq 1}$ suite de vecteurs L^2 de moyenne m, de matrice de covariance K non dégénérée. En posant $S^{(n)}$ la somme des n premiers termes de X, alors

$$\sqrt{K}^{-1}\frac{S^(n)-nm}{\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$$

et

$$\frac{S^(n)-nm}{\sqrt{n}} \Rightarrow \mathcal{N}(0,K).$$

Démonstration. Par le théorème de Lévy, on montre que $\varphi_{\sqrt{K}^{-1}\frac{S^(n)-nm}{\sqrt{n}}}(x) \to \exp\left(-\frac{\|x\|^2}{2}\right)$. Or pour $x\in\mathbb{R}$, $\varphi_{\sqrt{K}^{-1}\frac{S^(n)-nm}{\sqrt{n}}}(x) = \varphi_{\left\langle \sqrt{K}^{-1}\frac{S^(n)-nm}{\sqrt{n}},x\right\rangle}(1)$ et $\left\langle \sqrt{K}^{-1}\frac{S^(n)-nm}{\sqrt{n}},x\right\rangle = \frac{\left\langle S^{(n)},\left(\sqrt{K}^{-1}\right)^T x-n\langle\rangle\right\rangle}{\sqrt{n}}$ et le TCL usuel donne ce qu'il faut en considérant les $Y_k=\langle X^{(k)},y\rangle$. □

Théorème 10.0.4 (Fisher)**.** Soient $X_1,\dots,X_n$ des va iid $\mathcal{N}(m,\sigma^2)$ avec $\sigma>0$, alors avec

$$\overline{X_n}=\frac{1}{n}\sum_{i=1}^n X_i, \qquad S_n^2=\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X_n})^2$$

on a $\overline{X_n}\sim\mathcal{N}(m,\sigma^2/n) \perp\!\!\!\perp \frac{nS_n^2}{\sigma^2}\sim\chi^2(n-1)$ et $\sqrt{n-1}\frac{\overline{X_n}-m}{\sqrt{S_n^2}}\sim\mathcal{T}(n-1)$.

Rappel 10.0.5. Si $X\sim\mathcal{N}(0,1)\perp\!\!\!\perp Y\sim\mathcal{X}^2(n)$, $\frac{X}{\sqrt{\frac{Y}{n}}}\sim\mathcal{T}(n)$ suit la loi de Student.

Démonstration. Quitte à poser $X_i'=\frac{X-1-m}{\sigma}$, il nous suffit de montrer le résultat pour $m=0$ et $\sigma=1$.
Soit $u_1=\frac{1}{\sqrt{n}}(1,\dots,1)^T$ de norme 1, on complète u_1 en bon u et on note A la matrice dont les colonnes sont les u_i. A est orthogonale et pour $X\sim\mathcal{N}(0,I_d)$, $Y=A^TX\sim\mathcal{N}(0,AA^T)=\mathcal{N}(0,I_d)$. Si on pose $Y_1=\langle u_1,X\rangle=\frac{\sum_{i=1}^n X_i}{\sqrt{n}}$ ie $\overline{X_n}=\frac{Y_i}{\sqrt{n}}\sim\mathcal{N}\left(0,\frac{1}{n}\right)$.
Comme A est orthogonal $\|X\|^2=\|Y\|^2$ et $S_n^2=\sum_{i=1}^n(X_i-\overline{X_n})^2=\sum_{k=1}^n X_k^2-n\overline{X_n}^2=\sum_{k=2}^n Y_k^2\sim\chi^2(n-1)\perp\!\!\!\perp \frac{Y_1}{\sqrt{n}}=\overline{X_n}$. □

Théorème 10.0.6 (de Cochran)**.** Soit X un vecteur gaussien $\mathcal{N}(0,\sigma^2I_d)$ de $\mathbb{R}^d$. On considère $\mathbb{R}^d=V_1\oplus\dots\oplus V_k$ avec $\dim V_i=d_i$ et $V_i\perp\!\!\!\perp V_j$. Alors les $P_{V_i}(X)$ sont des vecteurs gaussiens indépendants et $\|P_{V_i}(X)\|^2\sim\chi^2(d_i)$

Exemple 10.0.7. Si $\mathbb{R}^d = F \oplus F^\perp$, $\|X\|^2 = \|P_F(X)\|^2 + \|P_{F^\perp}\|^2$ et les deux termes sont indépendents.

Trouver preive. □

Modèle linéaire gaussien

On suppose qu'on oberve les images bruitées de certaines données x. Étant donnés $(x_1, \dots, x_n)$, on observe $Y_i = \alpha + \beta x_i + E_i$ où les E_i sont iid de $\mathcal{N}(0,1)$ idd et σ est inconnu.
Si on note $\mathbf{1} = (1, \dots, 1)^T$, $E = (E_1, \dots, E_n)$ et $x = (x_1, \dots, x_n)$, on a $Y = \alpha\mathbb{1} + \beta x + \sigma E$. On suppose de plus que $F = \text{Vect}(\mathbb{1}, x)$ est de dimension 2. Pour $a \in \mathbb{R}^n$, on note $\overline{a} = \left\langle a, \frac{\mathbb{1}}{\|\mathbb{1}\|^2} \right\rangle$ la moyenne des a_i.

[cf cahier pour suite]

Chapitre 11

13/11/2023 - Géométrie Affine – Jérémy Le Borgne

"Une géométrie" : un ensemble (de points) et un groupe (de transformations) agissant sur un ensemble.
Pour l'agrégation :
- ce point de vue
- géométrie euclidienne "élémentaire"
- théorèmes "projectifs" (Desargues, Pappus,...)

11.1 Espaces affines

Soit K un corps.

11.1.1 Définitions

Définition 11.1.1. Un **espace affine** sur K est la donnée de :
- un ensemble non vide $\mathcal{E}$,
- un K-espace vectoriel E
- une action simplement transitive de E sur $\mathcal{E}$.

Dans ce cas, $\mathcal{E}$ est appelé l'**ensemble des points** et E l'**espace vectoriel directeur** (ou la **direction**).

Remarque 11.1.2. Être simplement transitive veut dire que l'action est transitive et libre (ie tous les stabilisateurs sont triviaux). Ce qui est différent de fidèle (l'intersection des stabilisateurs est triviale) même si libre implique fidèle mais la réciproque est fausse en général (eg considérer l'action de $\mathfrak{S}_3$ sur $\{1,2,3\}$ est transitive, fidèle mais pas libre. Mais dans le cas de l'action transitive d'un groupe abélien, fidèle et libre sont équivalents (car transitive implique que les stabilisateurs des points sont conjugués).

Exemples 11.1.3.
- Si E est un K-ev, le couple (E,E) avec l'action naturelle par translation est un espace affine.
- Si E est un K-ev et $F \subset E$ sous-espace vectoriel et $A \in E$. Soit $\mathcal{F} = A + F$. Alors $(\mathcal{F}, F)$ muni de (la restriction à $\mathcal{F}$ de) l'action naturelle est un espace affine.

Définition 11.1.4. Soit $(\mathcal{E}, E)$ un espace affine. Soient $A, B \in \mathcal{E}$. Par définition, il existe un unique $v \in E$ tel que $v \cdot A = B$ (puisque l'action est simplement transitive). Cet unique v est noté $\overrightarrow{AB}$. On notera plutôt l'action sous forme additive ie $A + \overrightarrow{AB} = B$.

Lemme 11.1.5. L'application $(\mathcal{E}, \mathcal{E}) \ni (A, B) \mapsto \overrightarrow{AB} \in E$ est surjective.

Démonstration. Soient $v \in E$ alors pour tout $A \in \mathcal{E}$ et $B = A + v$, $\overrightarrow{AB} = v$. Ce qui donne la surjectivité. □

Remarque 11.1.6. Elle n'est pas injective (car pour tout $A \in \mathcal{E}$, $\overrightarrow{AA} = 0$).

Lemme 11.1.7. Soit $(\mathcal{E}, E)$ un espace affine, soit $A_0 \in \mathcal{E}$. Alors $E \ni v \mapsto A_0 + v \in \mathcal{E}$ est une bijection de réciproque $A \mapsto \overrightarrow{A_0 A}$.

Démonstration. Pour $v \in E$ et $B = A_0 + v$, $\overrightarrow{A_0B} = v$ et pour $v, v' \in E$ tels que $B = A_0 + v = A_0 + v'$, alors $\overrightarrow{A_0B} = v = v'$. □

Conséquence 11.1.8. Un espace affine est toujours en bijection avec sa direction et chaque choix de point de $\mathcal{E}$ fournit une telle bijection.

Proposition 11.1.9 (Relation de Chasles)**.** Pour $(\mathcal{E}, E)$ un espace affine et $A, B, C \in \mathcal{E}$, alors $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$.

Démonstration. On a $C = B + \overrightarrow{BC} = A + \overrightarrow{AB} + \overrightarrow{BC}$, d'où le résultat. □

11.1.2 Sous-espaces affines

Définition 11.1.10. Soit $(\mathcal{E}, E)$ espace affine, soit $\mathcal{F} \subset \mathcal{E}$ non vide, alors $\mathcal{F}$ est un **sous-espace affine** si et seulement s'il existe $F \subset E$ sous-espace vectoriel de E tel que la restriction à $\mathcal{F}$ de l'action de F sur $\mathcal{E}$ munit $(\mathcal{F}, F)$ d'une structure d'espace affine.

Lemme 11.1.11. Pour $\mathcal{F} \subset \mathcal{E}$ non vide, on suppose qu'il existe $A_0 \in \mathcal{F}$ tel que $F = \left\{\overrightarrow{A_0A}, A \in \mathcal{F}\right\}$ soit un sous-espace vectoriel de E. Alors $\mathcal{F}$ est un sous-espace affine de $\mathcal{E}$.

Démonstration. Sous ces hypothèses, l'action de f sur $\mathcal{F}$ est transitive. Elle est libre comme restriction d'une action libre. Donc $\mathcal{F}$ est un sous-espace affine. □

Remarque 11.1.12. Si $\mathcal{F}$ sous-espace affine de $\mathcal{E}$ de direction $F \subset E$ et si $A_0 \in \mathcal{F}$ alors $\left\{\overrightarrow{A_0A}, A \in \mathcal{F}\right\} = F$. Le lemme donne ainsi une caractérisation des sous-espaces affines et le F obtenu ne dépend pas de A_0.

Proposition 11.1.13. Les sous-espaces affines de $(\mathcal{E}, E)$ sont les parties de la forme $A_0 + F$ avec $A_0 \in \mathcal{E}$ et F sous-espaces vectoriels de E.

Lemme 11.1.14. Toute intersection non vide de sous-espaces affines est un sous-espace affine.

à faire. □

Définition 11.1.15 (Sous-espace affine engendré par une partie)**.** Soit $(\mathcal{E}, E)$ un espace affine, soit $\mathcal{X} \subset \mathcal{E}$ non vide, alors on appelle **sous-espace affine engendré** par $\mathcal{X}$ le plus petit sous-espace affine contenant $\mathcal{X}$.

Remarque 11.1.16. Si $\mathcal{F} \subset \mathcal{E}$ est un sous-espace affine de direction F, alors $F = \left\{\overrightarrow{AB}, A \in \mathcal{F}, B \in \mathcal{F}\right\}$ mais cela ne caractérise pas les sous-espaces affines (eg les demi-droites).

Définition 11.1.17. On appelle **dimension d'un espace affine** la dimension de sa direction.

Exemples 11.1.18. On parle de point affine, de droite affine, de plan affine, d'hyperplan affine,...

Définition 11.1.19. On dit que deux sous-espaces affines sont dits **parallèles** si et seulement s'ils ont la même direction.

Remarque 11.1.20. On peut en fait définir la notion de parallélisme de manière asymétrique "$\mathcal{F}$ est parallèle à $\mathcal{F}'$" ssi $F \subset F'$.

Proposition 11.1.21. Deux droites parallèles sont confondues ou disjointes.

Proposition 11.1.22. Deux droites d'un plan affine non parallèles se rencontrent en exactement 1 point.

Démonstration. On note $(\mathcal{D}_1, D_1)$ et $(\mathcal{D}_2, D_2)$ ces deux droites. Par hypothèse $D_1 \neq D_2$ donc $E = D_1 \oplus D_2$ ainsi $\mathcal{D}_1 \cap \mathcal{D}_2$ contient au plus 1 point. Soient $A_1 \in \mathcal{D}_1$ et $A_2 \in \mathcal{D}_2$, $\overrightarrow{A_1A_2} \in E$ s'écrit $\overrightarrow{A_1A_2} = v_1 + v_2$ avec des notations évidentes. Donc $A_1 + v_1 + v_2 = A_2$ d'où $A_1 + v_1 = A_2 - v_2$ donc $\mathcal{D}_1 \cap \mathcal{D}_2 \neq \emptyset$. □

Proposition 11.1.23. Par un point d'un espace affine, il passe exactement un sea tel que ce sous-espace est un sous-espace donné soient parallèles.

11.1.3 Repères affines

Soit $(\mathcal{E}, E)$ un espace affine.

Définition 11.1.24. Soit $\{A_i, i \in I\} \subset \mathcal{E}$, on dit que $\{A_i, i \in I\}$ est un **repère affine** si et seulement s'il existe $i_0 \in I$ tel que $\left\{\overrightarrow{A_{i_0}A_i}, i \in I \setminus \{i_0\}\right\}$ soit une base de E.

Par relation de Chasles, c'est vrai en un $i_0 \in I$ si et seulement si c'est vrai pour tout $i \in I$.
Si $\dim \mathcal{E} = n$, tout repère affine est une partie à $n+1$ éléments de $\mathcal{E}$.

Définition 11.1.25. $(A_i)_{i\in I}$ est dite **affinement libre** si et seulement s'il existe $i_0 \in I$ tel que $\left\{\overrightarrow{A_{i_0}A_i}, i \in I \setminus \{i_0\}\right\}$ soit libre.

Proposition 11.1.26. Supposons que $\dim \mathcal{E} = n$, soit $(A_0, \dots, A_n) \in \mathcal{E}^{n+1}$. Alors cette famille est un repère affine si et seulement si l'une des deux conditions suivantes est vérifiée :
- $(A_0, \dots, A_n)$ est affinement libre ;
- le sous-espace affine engendré par $(A_0, \dots, A_n)$ est $\mathcal{E}$.

Lemme 11.1.27. Soit $A_0 \in \mathcal{X} \subset \mathcal{E}$, alors le sous-espace affine engendré par $\mathcal{X}$ a pour direction $X = \text{Vect}\left\{\overrightarrow{A_0A}, A \in \mathcal{X}\right\}$.

Démonstration. Laissée en exercice. [à faire] □

11.2 Applications affines

Définition 11.2.1. Soient $(\mathcal{E}, E)$ et $(\mathcal{E}', E')$ deux espaces affines. Alors $f : \mathcal{E} \to \mathcal{E}'$ est dite **affine** si et seulement s'il existe $A_0 \in \mathcal{E}$ et $g \in \mathcal{L}(E, E')$ tels que pour $A \in E$,

$$f(A) = f(A_0) + g(\overrightarrow{A_0A}) \quad \text{ie} \quad \overrightarrow{f(A_0)f(A)} = g(\overrightarrow{A_0A}).$$

Proposition 11.2.2. Si $f : \mathcal{E} \to \mathcal{E}'$ est affine, alors l'application $E \ni v \mapsto \overrightarrow{f(A_0)f(A)}$ ne dépend pas de A_0.

À faire. □

L'application linéaire ainsi obtenue s'appelle la partie linéaire ou application linéaire sous-jacente de f noté $\vec{f}$.
On note $\text{Aff}(\mathcal{E}, \mathcal{E}')$ l'ensemble des applications affines de $\mathcal{E}$ vers $\mathcal{E}'$.

Remarques 11.2.3. Si X est un ensemble et $\mathcal{E}$ un espace affine, l'ensemble des applications $\mathcal{F}(X, \mathcal{E})$ est un espace affine de direction $\mathcal{F}(X, E)$. En particulier, $\mathcal{F}(\mathcal{E}, \mathcal{E}')$ est un espace affine de direction $\mathcal{F}(\mathcal{E}, E')$ dont $\text{Aff}(\mathcal{E}, \mathcal{E}')$ est un sous-espace affine. Soient $f_1, f_2 \in \text{Aff}(\mathcal{E}, \mathcal{E}')$ affines. Si on pose $\overrightarrow{f_1f_2}(A) = \overrightarrow{f_1(A)f_2(A)}$, $\overrightarrow{f_1f_2} \in \text{Aff}(\mathcal{E}, E')$ de direction $\vec{f_2} - \vec{f_1}$. Finalement $\text{Aff}(\mathcal{E}, \mathcal{E}')$ est un espace affine de direction $\text{Aff}(\mathcal{E}, E')$.

L'application $\text{Aff}(\mathcal{E}, \mathcal{E}') \ni f \mapsto \vec{f} \in \mathcal{L}(E, E')$ est une application affine.

Proposition 11.2.4. $f : \mathcal{E} \to \mathcal{E}'$ affine est injective (resp. surjective, bijectif) ssi $\vec{f}$ l'est.

à faire. □

Quelques transformations affines remarquables : les translations ($\vec{f} = \text{id}$), les homothéties (de centre A_0 et de rapport λ) et les similitudes.

Théorème 11.2.5 (Théorème de Thalès)**.** Soient $\mathcal{D}$ et $\mathcal{D}'$ deux droites affines parallèles distinctes dans un plan affine. Soit $A \notin \mathcal{D}, \mathcal{D}'$. Soient $B, C \in \mathcal{D}$. Alors, (AB) rencontre $\mathcal{D}'$ en B' et (AC) rencontre $\mathcal{D}'$ en C'. Alors il existe une unique homothétie de centre A qui envoie B sur B' et C sur C'. En particulier, il existe un unique $\lambda \in K$ tel que

$$\overrightarrow{AB'} = \lambda\overrightarrow{AB} \; ; \quad \overrightarrow{AC'} = \lambda\overrightarrow{AB} \; ; \quad \overrightarrow{B'C'} = \lambda\overrightarrow{BC}.$$

Proposition 11.2.6. L'image d'un sea par une application affine et l'image réciproque d'un sea par une application affine sont des sea.

Proposition 11.2.7. Si $f : \mathcal{E} \to \mathcal{E}'$ est affine et $\mathcal{F}_1$ est parallèle à $\mathcal{F}_2$, alors $f(\mathcal{F}_1)$ est parallèle à $f(\mathcal{F}_2)$.

Démonstration. Laissées en exercice. □

Remarque 11.2.8. Préserver les sea ne caractérise pas les applications affines : la conjugaison complexe sur $\mathbb{C}^2$ $(z_1, z_2) \mapsto (\overline{z_1}, \overline{z_2})$ préserve les sea et n'est pas $\mathbb{C}$-affine (bien que $\mathbb{R}$-affine).

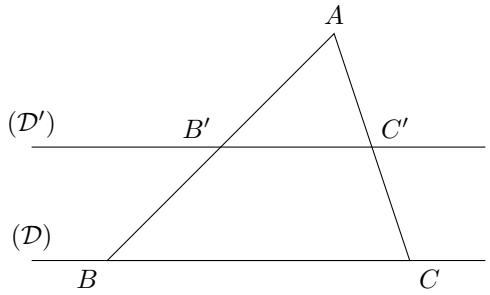

FIGURE 11.1 – Illustration du théoèreme de Thalès

11.2.1 Applications affines et repères affines

Proposition 11.2.9. Soient $(A_0, \dots, A_n)$ un repère affine de $\mathcal{E}$ et $B_0, \dots, B_n) \in \mathcal{E}'$. Alors il existe une unique $f : \mathcal{E} \to \mathcal{E}'$ affine telle que pour tout $i \in [\![0, n]\!]$.

Démonstration. Si on a une telle application f alors $f(A_0) = B_0$ et pour tout i, $\vec{f}(\overrightarrow{A_0A_i}) = \overrightarrow{B_0B_i}$. Comme $\overrightarrow{A_0A_i}$ est une base de E, il existe une unique application linéaire $g : E \to E'$ telle que $g(\overrightarrow{A_0A_i}) = \overrightarrow{B_0B_i}$.

Soit f l'application définie par $f(A) = B_0 + g(\overrightarrow{A_0A})$ alors pour tout i, $f(A_i) = B_i$ et f est affine. □

Proposition 11.2.10. Soit $f : \mathcal{E} \to \mathcal{E}'$ affine, alors f est bijective ssi l'image par f d'un repère affine est un repère affine.

Démonstration. Soit $(A_0, \dots, A_n)$ un repère affine et posons $B_i = f(A_i)$. Pour $i \geq 1$, $\overrightarrow{B_0B_i} = \vec{f}(\overrightarrow{A_0A_i})$. La famille $\{\overrightarrow{B}_0B_i, i \geq 1\}$ ssi $\vec{f}$ bijective ssi f bijective. □

Exemple 11.2.11. Dans le plan affine, les repères affines sont les triangles non aplatis et si ABC et $A'B'C'$ sont deux triangles non aplatis, il existe une unique application affine f telle que $A \mapsto A'$, $B \mapsto B'$ et $C \mapsto C'$.

Principe de démonstration. Si on a un énoncé dont les hypothèses et conclusions sont invariantes par transformation affine, alors il suffit de démontrer un cas particulier pour l'obtenir sur toute "l'orbite du cas particulier".

Exemple 11.2.12.

Théorème 11.2.13 (de Steiner)**.** Soit T un triangle d'un plan affine, alors il existe une unique ellipse tangente aux côtés de T en leurs milieux.

Trouver et faire preuve. □

11.2.2 Groupes affines

Soit $\mathcal{E}$ un espace affine, on note $GA(\mathcal{E}) = \{f \in \mathrm{Aff}(\mathcal{E}) \text{ bijective}\}$.

Soit $\Phi : f \in \mathrm{Aff}(\mathcal{E}) \to \mathcal{L}(E)$, Φ est un morphisme de groupes et alors $GA(\mathcal{E})$ est l'image réciproque par Φ de $GL(E)$. On note alors $T(\mathcal{E}) = \ker \Phi$ le groupe des translations et donc $T(\mathcal{E}) \triangleleft GA(\mathcal{E})$.

Si t_v est la translation de vecteur v et si $g \in GA(\mathcal{E})$, $gt_vg^{-1} = t_{\vec{g}(v)}$.

Principe. f et gfg^{-1} sont de même nature géométrique et les éléments caractéristiques de gfg^{-1} sont les images par g de ceux de f :

$$1 \to E \overset{groupes}{\simeq} T(\mathcal{E}) \to GA(\mathcal{E}) \overset{\Phi}{\to} GL(E) \to 1.$$

Cette suite est scindée.

Soit $A_0 \in \mathcal{E}$, $A \mapsto \overrightarrow{A_0A}$ est bijective de réciproque $v \mapsto A_0 + v$.

L'application $\psi : GL(E) \ni \vec{f} \mapsto (f : A \mapsto A_0 + \vec{f}(\overrightarrow{A_0A})) \in GA(\mathcal{E})$ est une section de Φ (ie $\Phi \circ \psi = \mathrm{id}$ et ψ morphisme de groupes).

Remarque 11.2.14. ψ réalise un isomorphisme de $GL(E)$ sur $GL(\mathcal{E})$ où $\mathcal{E}$ est muni de la structure d'espace vectoriel par $v \mapsto A_0 + v$ où A_0 est l'origine. $GA(\mathcal{E})$ est un produit semi-direct $GL(E) \ltimes E$.

11.2.3 Sous-groupe $HT(\mathcal{E})$

On note $HT(\mathcal{E}) = \Phi^{-1}(Z(GL(E)))$ le groupes des homothéties et translations. Alors $HT(\mathcal{E}) \triangleleft GA(\mathcal{E})$ et on a également $HT(\mathcal{E}) = \ker(\overline{\Phi}^{-1} : GA(E) \to PGL(E))$.

Remarques 11.2.15. — Si h est une homothétie de centre A et t_v est la translation de vecteur v alors $t_v h t_v^{-1}$ est une homothétie de même rapport et de centre $A + v$. ht_v est également une homothétie de même rapport que h et de centre $A + \frac{\lambda}{1-\lambda}v$ (cf. dessin).

[Faire dessin illustrant la remarque]

Remarque 11.2.16. $t_v h = h t_v$ ssi $\vec{h} = \mathrm{id}$ ou $v = 0$.

Proposition 11.2.17. $HT(\mathcal{E})$ est l'ensemble des éléments de $G1(\mathcal{E})$ qui envoient toute droite sur une droite parallèle.

Démonstration. □

Proposition 11.2.18. Soit $f \in GA(\mathcal{E})$, alors f a un unique point fixe ssi $\vec{f} - \mathrm{id}$ est injective.

Démonstration. Si f a un point fixe $A_0 \in \mathcal{E}$, alors pour tout $A \in \mathcal{E}$,

$$f(A) = A \Leftrightarrow f(A_0) + \vec{f}(\overrightarrow{A_0A}) = A_0 + \overrightarrow{A_0A} \Leftrightarrow \vec{f}(\overrightarrow{A_0A}) = \overrightarrow{A_0A}.$$

Donc, dans ce cas, l'ensemble des points fixes de f est un sea de direction $\ker(\vec{f} - \mathrm{id})$.

Réciproquement, supposons $\ker(\vec{f} - \mathrm{id}) = \{0\}$.
Montrons que f admet un point fixe.
Soit $A_0 \in \mathcal{E}$, soit $A \in \mathcal{E}$,

$$f(A) = A \Leftrightarrow f(A_0) + \vec{f}(\overrightarrow{A_0A}) = A_0 + \overrightarrow{A_0A} \Leftrightarrow \overrightarrow{A_0f(A_0)} = (\vec{f} - \mathrm{id})(\overrightarrow{A_0A}).$$

D'après le théorème du rang, $\mathrm{Im}(\vec{f} - \mathrm{id}) = E$. En particulier, il existe v tel que $\overrightarrow{A_0f(A_0)} = (\vec{f} - \mathrm{id})(v)$. Soit $A = A_0 + v$, alors $f(A) = A$. □

Théorème 11.2.19 (de décomposition)**.** Soit f affine, on suppose que $E = \ker(\vec{f} - \mathrm{id}) \oplus \mathrm{Im}(\vec{f} - \mathrm{id})$. Alors il existe un unique couple $(v, g) \in E \times \mathrm{Aff}(\mathcal{E})$ tel que g ait un point fixe et $f = t_v g = g t_v$.

Démonstration. [Essayer de rédiger l'analyse-synthèse correctement]
On a $t_v g = g t_v$ ssi $v \in \ker(\vec{g} - \mathrm{id})$. Par ailleurs, examinons quand $t_{-v} f = t_v^{-v} f$ a un point fixe. Soit $A \in \mathcal{E}$, on note $B = f(A)$. $t_{-v}f$ a un point fixe ssi $\overrightarrow{A}(t_{-v}f(A)) = \overrightarrow{AB} - v \in \mathrm{Im}(\vec{f} - \mathrm{id})$ (et $\vec{f} = \overrightarrow{t_{-v}f}$). On peut écrire $f = t_v g$ tel que t_v et g commutent et g a un point fixe dès que $v \in \ker(\vec{f} - \mathrm{id})$ et $\overrightarrow{AB} - v \in \mathrm{Im}(\vec{f} - \mathrm{id})$. Soit $\overrightarrow{AB} = v + (\overrightarrow{AB} - v)$, la décomposition de $\overrightarrow{AB}$ sur la somme directe $\ker(\vec{f} - \mathrm{id}) \oplus \mathrm{Im}(\vec{f} - \mathrm{id})$. Par construction, $\overrightarrow{AB} - v \in \mathrm{Im}(\vec{f} - \mathrm{id})$ donc $g = t_{-v}f$ a un point fixe et $\vec{g} = \vec{f}$ donc $v \in \ker(\vec{g} - \mathrm{id})$ ie $t_v g = f = g t_v$.
Pour l'unicité, supposons que f possède une telle décomposition. Pour $A_0 \in \mathcal{E}$ un point fixe de g, $f(A_0) = g(A_0) + u = A_0 + u$ donc $\overrightarrow{A_0f(A_0)} = u$. Soit $A \in \mathcal{E}$, $A = A_0 + \overrightarrow{A_0A}$ donc

$$f(A) = \underbrace{f(A_0)}_{\in \ker(\vec{f} - \mathrm{id})} + \underbrace{\vec{f}(\overrightarrow{A_0A})}_{\in \mathcal{I}\mathrm{m}(\vec{f} - \mathrm{id})}.$$

La composante de $\overrightarrow{Af(A)}$ selon $\ker(\vec{f} - \mathrm{id})$ (dans la décomposition $E = \ker(\vec{f} - \mathrm{id}) \oplus \mathrm{Im}(\vec{f} - \mathrm{id})$) est indépendante de A et doit être égale à $\vec{u}$. D'où l'unicité de $\vec{u}$ et de g. □

Remarque 11.2.20. Deux cas particulier où l'hypothèse $E = \ker(\vec{f} - \mathrm{id}) \oplus \mathrm{Im}(\vec{f} - \mathrm{id})$ sont les cas $\vec{f}$ est diagonalisable et E espace euclidien et $\vec{f} \in \mathcal{O}(E)$, $GAE(\mathcal{E}) = \{f \in GA(\mathcal{E}) : \vec{f} \in \mathcal{O}(E)\}$.
Dans le deuxième cas, $\ker(\vec{f} - \mathrm{id}) \perp \mathrm{Im}(\vec{f} - \mathrm{id})$. Soit $x \in \ker(\vec{f} - \mathrm{id})$ et soit $z \in E$, $y = \vec{f}(z) - z$, alors

$$\langle x, y \rangle = \langle x, \vec{f}(z) - z \rangle = \langle x, \vec{f}(z) \rangle - \langle x, z \rangle = \langle x - \vec{f}(x), \vec{f}(z) \rangle + \langle \vec{f}(x), \vec{f}(z) \rangle - \langle x, z \rangle = 0.$$

11.3 Barycentres

Soit $\mathcal{E}$ espace affine sur K, soient $\lambda_0, \dots, \lambda_n \in K$ et $A_0, \dots, A_n \in \mathcal{E}$. L'application $\mathcal{E} \ni M \mapsto \sum_{i=0}^n \lambda_i \overrightarrow{A_i M}$ est affine et sa partie linéaire est l'homothétie de rapport $\sum_i \lambda_i$.

[Justifier]

Définition 11.3.1. Un système de points pondérés est une famille $\{(A_i, \lambda_i), 0 \leq i \leq n\}$ telle que $A_i \in \mathcal{E}$ où les $\lambda_i \in K$ et $\sum_i \lambda_i \neq 0$.

Définition 11.3.2. Le barycentre du système de points pondérés $(A_0, \lambda_0), \dots, (A_0, \lambda_n)$ est l'unique M tel que $\sum_{i=0}^n \lambda_i \overrightarrow{A_i M} = \vec{0}$.

Lemme 11.3.3. G est le barycentre du SPP $\{(A_i, \lambda_i), 0 \leq i \leq n\}$ tel que $\sum_i \lambda_i = 1$ ssi pour tout $M \in \mathcal{E}$, $h(M) = \overrightarrow{GM}$.

Faire preuve. □

Lemme 11.3.4. Le sea engendré par $\mathcal{X} \subset \mathcal{E}$ est l'ensemble des barycentres de SPP de points de $\mathcal{X}$.

Proposition 11.3.5. Les applications affines préservent les barycentres.

Faire preuve. □

Théorème 11.3.6. Les transformations qui préservent les barycentres sont affines.

Démonstration. Soit (A_i) un repère affine, soit f préservant les barycentres. Soit $M \in \mathcal{E}$,

$$\overrightarrow{A_0 M} = \sum_{i=1}^n \alpha_i \overrightarrow{A_0 A_i}$$

$$\text{d'où} \quad 0 = \sum_{i=1}^n \alpha_i \left(\overrightarrow{A_0 A_i} - \overrightarrow{A_0 M} \right) - \overrightarrow{A_0 M} + \left(\sum_{i=1}^n \alpha_i \right) \overrightarrow{A_0 M} = \sum_{i=1}^n \alpha_i \overrightarrow{M A_i} + \left(1 - \sum_{i=1}^n \alpha_i \right) \overrightarrow{M A_0}.$$

Ainsi

$$M = \text{Bar}\left(\left\{ (A_0, 1 - \sum_i \alpha_i), (A_1, \alpha_1), \cdots, (A_n, \alpha_n) \right\} \right)$$

et

$$f(M) = \text{Bar}\left(\left\{ (f(A_0), 1 - \sum_i \alpha_i), (f(A_1), \alpha_1), \cdots, (f(A_n), \alpha_n) \right\} \right).$$

Soit g la transformation affine qui envoie A_i sur $f(A_i)$ pour tout i. Par hypothèse, f coïncide avec tous les barycentres de points de $(A_0, \dots, A_n)$ donc sur $\mathcal{E}$. □

11.3.1 Repère et coordonnées barycentriques

Définition 11.3.7. Un **repère barycentrique** est une famille $A_0, \dots, A_n \in \mathcal{E}$ telle que pour tout $M \in \mathcal{E}$, il existe des uniques $\lambda_0, \dots, \lambda_n \in K$ de somme 1 tels que $M = \text{Bar}(\{(A_0, \lambda_0), (A_1, \lambda_1), \cdots, (A_n, \lambda_n)\})$.

Remarque 11.3.8. Il y a une correspondance entre le cas affine et barycentrique :

$$M = A_0 + \sum_{i=1}^n \alpha_i \overrightarrow{A_0 A_i} \quad \Longleftrightarrow \quad M = \text{Bar}\left(\left\{ (A_0, 1 - \sum_i \alpha_i), (A_1, \alpha_1), \cdots, (A_n, \alpha_n) \right\} \right).$$

Définition 11.3.9. Soient $A_0, \dots, A_n \in \mathcal{E}$ un repère barycentruqye de $\mathcal{E}$. On dit que $(\lambda_0, \dots, \lambda_n)$ est un système de coordonnées barycentriques de $M \in \mathcal{E}$ dans ce repère ssi $M = \text{Bar}(\{(A_0, \lambda_0), (A_1, \lambda_1), \cdots, (A_n, \lambda_n)\})$.

Exemple 11.3.10. Le centre de gravité du triangle ABC admet pour système de coordonnées barycentriques $(1, 1, 1)$ mais également $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

Soient $A_0, \dots, A_n$ un repère barycentrique, $M \in \mathcal{E}$, on note alors

$$\forall i \in [\![1, n]\!],\ \lambda_i = \det\left(\overrightarrow{MA_{i+1}}, \dots, \overrightarrow{MA_n}, \overrightarrow{MA_0}, \dots, \overrightarrow{MA_{i-1}}\right) \quad \text{et} \quad \lambda_0 = 1 - \sum_{i=1}^{n} \lambda_i.$$

Les λ_i forment un système de coordonnées barycentriques de M.

Dans le cas du triangle ABC, on obtient la chose suivante : $\lambda_A, \lambda_B, \lambda_C = \left(\det\left(\overrightarrow{MB}, \overrightarrow{MC}\right), \det\left(\overrightarrow{MC}, \overrightarrow{MA}\right), \det\left(\overrightarrow{MA}, \overrightarrow{MB}\right)\right.$,

[Triangle ABC avec point M tq on dessine les sous-triangles AMC, CMB et BCA dont les aires sont définies par les λ_i]

Montrons que $\lambda_A\overrightarrow{MA} + \lambda_B\overrightarrow{MB} + \lambda_C\overrightarrow{MC} = 0$ ie

$$\det\left(\overrightarrow{MB}, \overrightarrow{MC}\right)\overrightarrow{MA} + \det\left(\overrightarrow{MC}, \overrightarrow{MA}\right)\overrightarrow{MB} + \det\left(\overrightarrow{MA}, \overrightarrow{MB}\right)\overrightarrow{MC} = 0.$$

Si on note $u = \overrightarrow{MA}$, $v = \overrightarrow{MB}$ et $w = \overrightarrow{MC}$, on cherche à montrer que $\det(u, v)u + \det(w, u)v + \det(u, v)w = 0$.

$(u, v, w) \mapsto \det(u, v)u + \det(w, u)v + \det(u, v)w$ est une application trilinéaire alternée donc proportionnelle au déterminant. Comme $\left(\overrightarrow{MA}, \overrightarrow{MB}, \overrightarrow{MC}\right)$ est liée, l'expression prise en ces 3 vecteurs vaut bien 0.

[faire diagramme triangulaire de E^3 dans V par f, de E^3 dans $\bigwedge^3 E$ (de dimension 1) par le produit vectoriel et de $\bigwedge^3 E$ dans V par $\bar{f}$]

[triangle cf .svg]

Exemple 11.3.11. $D = \mathrm{Bar}((A, 1), (B, 1), (C, -1))$

[calcul + dessin]

Exemple 11.3.12. Si $\lambda \neq 1$ et $A \in \mathcal{E}$, on pose $h : M \mapsto Bar((A, 1 - \lambda), (M, \lambda))$ alors h est l'homothétie de centre A et de rapport λ

11.3.2 Théorème fondamental de la géométrie affine

Définition 11.3.13. Soit $\sigma \in \mathrm{Aut}(K)$ automorphisme de corps et soit E un K-ev. On dit que $f : E \to E$ est semi-linéaire ssi pour tous $x, y \in E$, $\lambda \in K$, $f(x + \lambda y) = f(x) + \sigma(x)f(y)$.

Exemples 11.3.14.
- Si $\sigma = \mathrm{id}$, σ-semi-linéaire = linéaire.
- Si $K = \mathbb{C}$ et $\sigma(z) = \bar{z}$, $f : (z_1, \dots, z_n) \mapsto (\bar{z}_1, \dots, \bar{z}_n)$ est σ-semi-linéaire.
- Si $K = \mathbb{R}$, σ vaut nécessairement id.

Définition 11.3.15. Une application $f : E \to E$ est dite semi-linéaire ssi il existe $\sigma \in \mathrm{Aut}(K)$ tq f σ-semi-linéaire.

Remarque 11.3.16. Sur $\mathbb{R}$, semi-linéaire <=> linéaire

Théorème 11.3.17. Soit $\mathcal{E}$ un espace affine de dimension au moins 2 sur $K \neq \mathbb{F}_2$. Alors, si $f : \mathcal{E} \to \mathcal{E}$ est bijective et préserve l'alignement, alors f est semi-affine.

Remarques 11.3.18. Ce théorème est faux en dimension 1 : toute bijection préserve l'alignement.

Cela n'a pas de sens sur $\mathbb{F}_2$ car les droites n'ont que 2 points.

Démonstration. Soit $f : \mathcal{E} \to \mathcal{E}$ qui préserve l'alignement ie f envoie toute droite sur une droite. [TODO : faire dessin] Ainsi $f((AB)) = (f(A)f(B))$.

f préserve les sea et leurs dimensions. Soit $\mathcal{F}$ sea de $\mathcal{E}$ de repère affine $A_0, \dots, A_n$. On note $B_i = f(A_i)$ pour tout i. Montrons que $f(\mathcal{F})$ est le sea engendré par les B_i. Soit $M \in \mathcal{F}$, montrons que $f(M)$ est dans ce sous-espace affine. On procède par récurrence sur $\dim F$.....

Montrons que f préserve le parallélisme. Soient $A_0 \in \mathcal{E}$, $u, v \in E$. Soit $(A_0, A_0 + u, A_0 + v, A_0 + u + v)$ un parallélogramme de centre I [TODO : faire dessin]

$(f(A_0)f(I))$ passe par $f(A_0 + u + v)$ [TODO : dessin mais faire un quadrilatère qcq] □

Chapitre 12

16/11/2023 - Espérance conditionnelle – Jürgen Angst

[cf. notes]

Chapitre 13

20/11/2023 - Géométrie projective – Jérémy Le Borgne

Soit E un K-ev de dim au moins 1 et $P(E) = E \setminus \{0\}/K^\times = \{\text{droites de } E\}$. L'action de $GL(E)$ sur E induit une action sur $P(E)$ et $Z(GL(E))$ agit trivialement sur $P(E)$ donc $PGL(E)$ agit naturellement sur $P(E)$. On appelle homographies les éléments de $PGL(E)$.

On considère $E = K^{n+1}$ pour $n \geq 0$.

Définition 13.0.1. Si $x = (x_0, \ldots, x_n) \in K^{n+1}$, on dit que $[x_0 : \cdots : x_n]$ est un système de coordonnées homogènes de $\bar{x}$ (image de x dans $P(E)$).

Remarque 13.0.2. Pour $\lambda \neq 0$, $[\lambda x_0 : \cdots : \lambda x_n] = [x_0 : \cdots : x_n]$.

Ainsi

$$P(K^{n+1}) = \{\bar{x} \in P(E) : x_n = 0\} \sqcup \{\bar{x} \in P(E) : x_n \neq 0\} = P(K^n) \sqcup K^n$$

où on identifie $P(K^n)$ à l'image de $[x_0 : \cdots : x_{n-1} : 0] \mapsto [x_0 : \cdots : x_{n-1}]$ et K^n à l'image de $[x_0 : \cdots : x_{n-1} : \underset{\neq 0}{x_n}] \mapsto (x_0/x_n, \cdots, x_{n-1}/x_n)$.

Plus intrinsèquement,

.

Théorème 13.0.3. Dans le plan projectif, 2 droites distinctes se rencontrent en exactement un point.

Bibliographie

[Per96] Daniel Perrin. *Cours d'algèbre*. Ellipses Paris, 1996.

[MM12] R. Mansuy et R. Mneimé. *Algèbre linéaire : Réduction des endomorphismes*. Vuibert, 2012.