

MATRICES ALÉATOIRES, GRAPHERS ALÉATOIRES ET QUELQUES APPLICATIONS AUX MATRICES DE CORRÉLATIONS

LAURE DUMAZ

RÉSUMÉ. Le point de départ de ce cours est un résultat remarquable d'un article génétique [6] publié dans la revue *Nature* en 2008 qui illustre le lien entre le génome d'individus européens et leur lieu d'origine. A l'aide de la technique de l'Analyse en Composante Principale, les auteurs ont produit une carte des variations génétiques qui ressemble de manière frappante à la carte de l'Europe. Nous allons analyser ce résultat à l'aide de la théorie des matrices aléatoires. En particulier, nous introduirons une méthode géométrique développée récemment pour étudier les matrices aléatoires.

Ce cours a été donné à l'occasion du colloque inter'action à Bordeaux en 2019. Attention : c'est un brouillon avec probablement beaucoup de typos/erreurs!

TABLE DES MATIÈRES

partie 1. Résultats de l'article et explications	2
1. Résultats de l'article "Genes mirror geography within Europe" [6]	2
2. Modélisation théorique avec des matrices aléatoires	2
2.1. Modélisation théorique du problème initial	2
2.2. Analyse en composantes principales (ACP)	4
2.3. Modèle de Wishart	7
partie 2. Matrices aléatoires	8
2.4. Introduction rapide aux principales grandes questions des matrices aléatoires	8
2.5. Transition de Baik, Ben-Arous, Pécché	11
partie 3. Matrices et graphes aléatoires	12
3. Trigonalisation et mesure spectrale	12
4. Convergence de graphe	14
partie 4. Application aux matrices de Wishart : Loi de Marchenko Pastur et transition de Baik, Ben-Arous, Pécché	17
5. Convergence de la plus grande valeur propre	18
6. Preuve de la loi de Marchenko-Pastur	19
7. Preuve partielle de la transition de Baik, Ben Arous, Pécché	20
8. Limites des processus et heuristique de convergence	21
Références	22

Première partie 1. Résultats de l'article et explications

1. RÉSULTATS DE L'ARTICLE "GENES MIRROR GEOGRAPHY WITHIN EUROPE" [6]

L'article "Genes mirror geography within Europe" publié dans Nature 2008 [6] étudie les variations génétiques d'environ 3000 individus européens à l'aide de centaines de milliers d'informations génétiques de leur ADN. Alors que les européens se différencient assez peu d'un point de vue génétique, l'article met en évidence une relation très forte entre les distances génétiques et les distances géographiques.

On considère $n \simeq 3000$ individus et $p \simeq 500000$ informations génétiques (500568 loci)¹ sur chacun de ces individus. Ces informations sont écrites dans une matrice X possédant n lignes et p colonnes : nous numérotions les individus $i = 1, \dots, n$ et les loci $\ell = 1, \dots, p$. Ainsi chaque ligne correspond aux divers loci d'un individu et l'élément $X_{i,\ell}$ correspond au ℓ -ième locus du i -ème individu. Pour simplifier, nous dirons que un locus est codé par un nombre réel. De plus on suppose qu'on a des données *centrées* c'est à dire que $\sum_{k=1}^p X_{i,k} = 0$. Cette hypothèse permet de comparer les individus entre eux. Par ailleurs, la location géographique de chaque individu est connue et donnée par le pays d'origine de ses grands-parents. Cette information n'est pas inscrite dans la matrice X . C'est une information supplémentaire.

On construit la *matrice de corrélation empirique* $C := \frac{1}{p} X^t X$ (matrice symétrique de dimension n). L'élément $C_{ij} := \frac{1}{p} \sum_{k=1}^p X_{ik} X_{jk}$ correspond à la "distance génétique" entre les individus i et j : elle mesure la corrélation empirique entre les individus i et j .

La matrice C est symétrique, elle est donc diagonalisable en base orthonormée. Traçons l'histogramme des valeurs propres. Si la matrice X de départ avait des entrées totalement aléatoires (i.i.d.), cet histogramme suivrait une loi bien précise, la loi de Marchenko-Pastur. On remarque ici que l'histogramme suit très précisément la loi de Marchenko-Pastur, sauf pour deux très grandes valeurs propres λ_1 et λ_2 . Notons φ_1 et φ_2 les deux vecteurs propres associés à ses deux valeurs propres. Ce sont des vecteurs de dimension n qui possèdent une entrée pour chaque individu. Pour chaque coordonnée k , on trace le point $(\varphi_1(k), \varphi_2(k))$ dans le plan, et on colorie le point en fonction de son lieu d'origine pour obtenir les figures 1 et 2. Le résultat frappant de cet article est que les deux premiers vecteurs propres (associés aux deux seules grandes valeurs propres) correspondent (à peu de modifications près) aux vecteurs longitude et latitude. Ainsi, l'ADN des européens est très informatif sur leur lieu d'origine : une carte de l'Europe est encodée dans leur gènes ! Grâce à cette analyse, les auteurs de l'article arrivent à placer 50% des individus à moins de 310kms de leur lieu d'origine et 90% à moins de 700kms. Ce résultat remarquable a bien-sûr de potentielles importantes applications médicales ou génétiques.

2. MODÉLISATION THÉORIQUE AVEC DES MATRICES ALÉATOIRES

2.1. Modélisation théorique du problème initial. Supposons que l'on ait n quantités que l'on appellera par la suite **individus** (mais qui peuvent être des personnes, des particules, des documents, des stocks de marché financier etc.) dont on veut comprendre la dynamique. On possède p informations sur chacun des individus (par exemple des informations génétiques dans notre cas, ou pour les stocks, leur rendement quotidien) notées

1. Un locus est une position d'un gène ou marqueur génétique sur un chromosome.

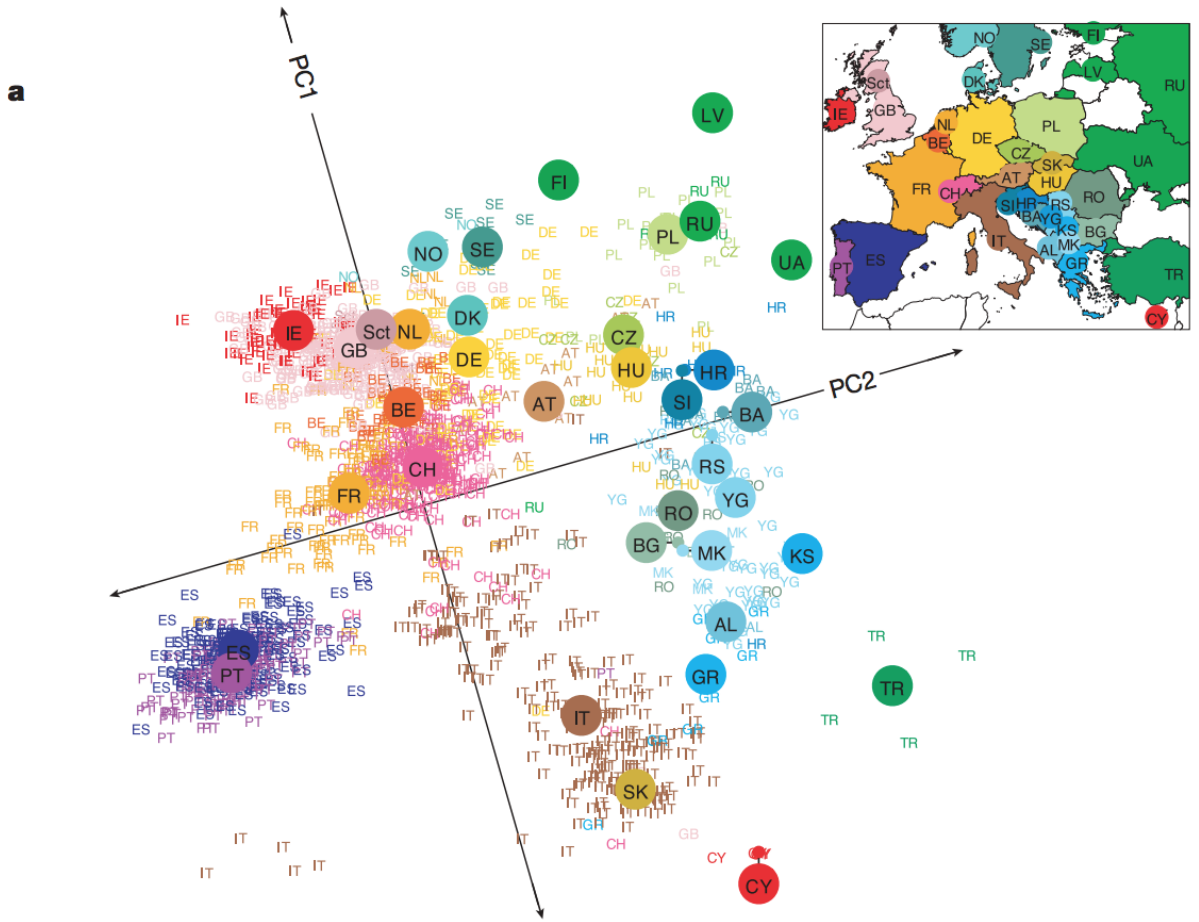


FIGURE 1. Projection des données génétiques de 1387 européens sur les axes de composantes principales PC1 et PC2. Les axes ont été pivotés par les auteurs pour souligner la similarité avec la carte de l’Europe.

$X_{i,k}$ pour l’information $k \in \{1, \dots, p\}$ sur l’individu $i \in \{1, \dots, n\}$. Les n individus sont tirés au hasard parmi une population beaucoup plus grande. Par conséquent, les données les concernant sont aléatoires (ce sont n individus “typiques”).

Pour simplifier, on suppose que les informations sont normalisées i.e. $\mathbb{E}[X_{ik}] = 0$. De plus on suppose que les vecteurs d’information sur les individus i.e. les colonnes de la matrice X notées $Y_k := (X_{1k}, \dots, X_{nk})$ pour $k = 1, \dots, p$ sont i.i.d. (mais on ne suppose pas que X_{ik} est indépendant ou de même loi que X_{jk} pour $j \neq i$).

La matrice de covariance ($\Sigma_{i,j} := \mathbb{E}[X_{ik}X_{jk}]$ qui mesure la “distance” entre les individus i et j est a priori inconnue (c’est la “vraie matrice de covariance”). Pour la mesurer à l’aide des données, on construit la matrice de corrélation empirique

$$C_{i,j} := \frac{1}{p} X^t X := \frac{1}{p} \sum_{k=1}^p X_{ik} X_{jk}$$

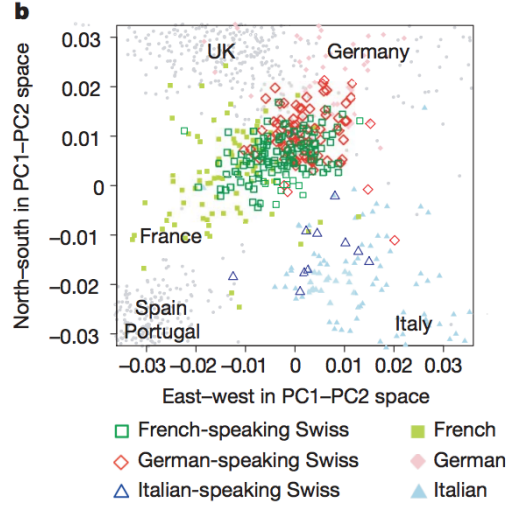


FIGURE 2. Zoom autour de la Suisse qui montre une différenciation entre les Suisses par leur langue. L'information contenue dans les gènes peut ainsi donner des informations géographiques très précises.

Lorsqu'on possède un grand nombre d'informations sur les individus i.e. dans la limite $p \rightarrow \infty$, la matrice C converge vers la matrice Σ grâce à la loi des grands nombres. Si on a par exemple $n = 4$ individus et $p = 10^6$ informations i.i.d, la matrice C donne a priori une bonne approximation de la matrice Σ .

Un autre régime classique est celui où p est grand mais n aussi avec n/p d'ordre de grandeur 1 i.e. $p \rightarrow \infty$ avec $n \rightarrow \infty$ et $n/p \rightarrow \gamma \in (0, +\infty)$. Il est intéressant de voir à quel point C diffère de Σ dans ce cas, par rapport au cas où $\gamma \rightarrow 0$. On va voir un peu plus loin que ce régime asymptotique possède aussi une universalité.

2.2. Analyse en composantes principales (ACP). La matrice C est une matrice symétrique positive. On peut donc décomposer C de la manière suivante :

$$C = OD^tO$$

avec D diagonale d'éléments les valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ et O matrice orthogonale (i.e. $O^tO = I$) de colonnes les vecteurs propres (normalisés) $(\varphi_1, \dots, \varphi_n)$. Cette décomposition est la clef de la *Analyse en composantes principales* (ACP).

Si l'on visualise les données comme un nuage de p points dans un espace de dimension n , la PCA permet de projeter ce nuage sur un espace de dimension plus petite, en utilisant seulement les premières composantes principales. Dans ce nouveau système de coordonnées, la plus grande variance des données correspond à la première coordonnée, la deuxième plus grande variance à la deuxième coordonnée etc.

Plus précisément, les Composantes Principales sont les vecteurs propres de C . La première composante principale est le vecteur propre φ_1 associé à la plus grande valeur propre

λ_1 . Ces quantités s'expriment de la façon suivante :

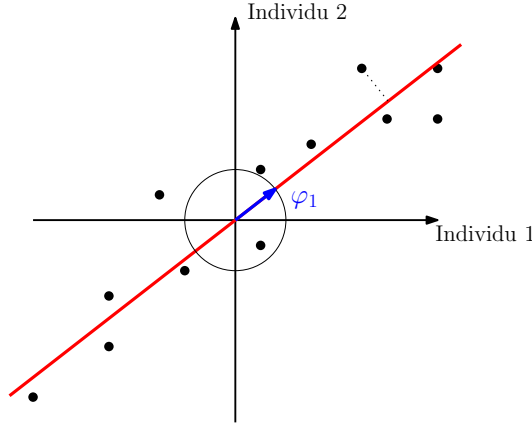
$$\lambda_1 := \max_{\varphi : \|\varphi\|=1} \langle C\varphi, \varphi \rangle = \max_{\varphi : \|\varphi\|=1} \sum_k \left(\sum_i X_{i,k} \varphi_i \right)^2,$$

$$\varphi_1 := \operatorname{argmax}_{\{\varphi : \|\varphi\|=1\}} \langle C\varphi, \varphi \rangle = \operatorname{argmax}_{\{\varphi : \|\varphi\|=1\}} \langle {}^tX\varphi, {}^tX\varphi \rangle .$$

Ainsi, φ_1 est la combinaison linéaire (qui met des poids sur les individus) de norme L^2 égale à 1 qui possède la plus grande *variance empirique*.

Un autre façon plus géométrique de visualiser le vecteur φ_1 est de procéder de la façon suivante : la quantité $\sum_{k=1}^p \|X_{\cdot,k}\|^2 - \langle {}^tX\varphi, {}^tX\varphi \rangle = \sum_{k=1}^p \|X_{\cdot,k}\|^2 - \sum_{k=1}^p \langle X_{\cdot,k}, \varphi \rangle^2$ mesure la somme des carrés des distances des points à la droite de direction φ . Ainsi le vecteur φ_1 est celui qui *minimise cette distance*.

Exemple : Supposons que l'on ait $n = 2$ individus/stocks de marché financier et $p = 12$ caractéristiques/rendements journaliers. Le vecteur φ_1 maximise la somme des projections



orthogonales sur la droite de direction φ_1 .

Les composantes suivantes sont définies de manière analogue itérativement en projetant sur le complément orthogonal des précédentes composantes. Ainsi,

$$\lambda_2 := \max_{\varphi \perp \varphi_1 : \|\varphi\|=1} \langle C\varphi, \varphi \rangle, \quad \varphi_2 := \operatorname{argmax}_{\{\varphi \perp \varphi_1 : \|\varphi\|=1\}} \langle C\varphi, \varphi \rangle .$$

l'ACP consiste ensuite à ne considérer que l'information provenant des N premiers vecteurs propres où $N < n$ et est souvent beaucoup plus petit que n c'est à dire à remplacer C par $\sum_{i=1}^N \lambda_i \varphi_i {}^t\varphi_i$.

L'ACP est particulièrement utile lorsqu'il y a une forte séparation entre les valeurs propres. Si par exemple λ_1 est beaucoup plus grande que λ_2 , on peut espérer que la corrélation entre les éléments considérés s'explique principalement par ce biais là. C'est le cas par exemple lorsque l'on considère des données financières, où φ_1 représente le marché.

Dans notre exemple, les deux premières valeurs propres sont prédominantes. La figure 1 correspond ainsi aux projections de chaque individu i représenté par le vecteur de la base canonique e_i sur les deux premiers vecteurs propres, autrement dit $\langle e_i, \varphi_1 \rangle$ et $\langle e_i, \varphi_2 \rangle$. Ainsi, le premier vecteur propre correspond à la longitude et le deuxième à la latitude.

Si les premières valeurs propres sont beaucoup plus grandes que les autres, on peut espérer que les vecteurs propres associés expliquent une grande partie des corrélations entre les individus, et que les autres ne sont dues qu'au bruit aléatoire. Cela n'est cependant pas tout à fait exact : on verra plus loin que les valeurs propres provenant du bruit peuvent être assez grandes ! La théorie des matrices aléatoires permet de prédire à quel point elles peuvent être grandes et donc de distinguer le bruit de la vraie corrélation.

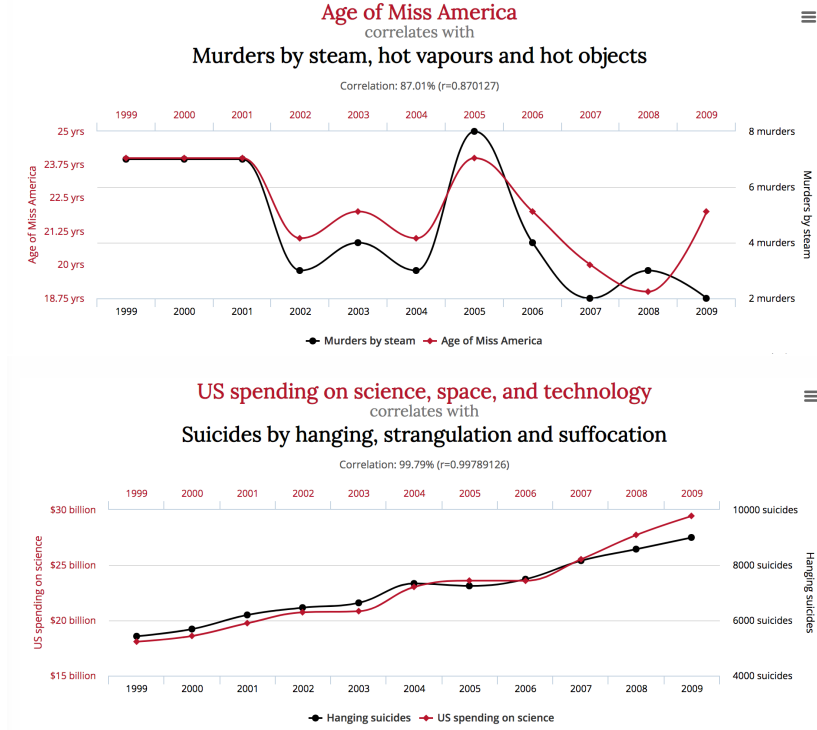


FIGURE 3. Fausses corrélations entre des variables fantaisistes. Source : <http://www.tylervigen.com/spurious-correlations>. Ce qui permet de distinguer une “fausse” corrélation d’une “vraie” est sa capacité prédictive.

Autre point de vue (probabiliste) : On peut appliquer l’ACP à la vraie matrice de covariance Σ (au lieu de la matrice de covariance empirique C) : $\Sigma = O'D^\Sigma tO'$, où $O' = (\phi_1, \dots, \phi_n)$ et D^Σ matrice diagonale de diagonale $\lambda_1^\Sigma > \dots > \lambda_n^\Sigma$, on a :

$$Y_1 = \sum_{\ell=1}^n \langle \phi_\ell, Y_1 \rangle \phi_\ell.$$

(Rappelons que Y_1 est la première colonne de la matrice X) Calculons la variance de $\langle \phi, Y_1 \rangle$:

$$\begin{aligned} \text{Var}(\langle \phi, Y_1 \rangle) &= \mathbb{E}[\langle \phi, Y_1 \rangle^2] \\ &= \mathbb{E}\left[\sum_{i,k} \phi(i)X_{i,1} \phi(k)X_{k,1}\right] \\ &= {}^t\phi \Sigma \phi \end{aligned}$$

Ainsi, sous la contrainte $\|\phi\| = 1$, en prenant $\phi = \phi_1$, on maximise la (vraie) variance que l'on peut obtenir, et cette variance vaut λ_1^Σ .

Ainsi, en décomposant les premiers vecteurs propres sur la base canonique, on obtient les combinaisons linéaires des coefficients de Y_1 (de norme L^2 égale à 1) qui expliquent le plus la variance de Y_1 .

2.3. Modèle de Wishart. Supposons dans ce paragraphe que $\Sigma = Id$ i.e. que toutes les entrées de la matrice X sont i.i.d de variance 1. Supposons de plus qu'elles sont de loi normale $\mathcal{N}(0, 1)$ (les individus sont indépendants). La matrice X est appelée *matrice de Wishart*. La densité de la loi jointe des valeurs propres de la matrice $C = \frac{1}{p} X^t X$ vaut :

$$f(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_{n,p}} \prod_{i=1}^n \lambda_i^{(p-n+1)/2-1} e^{-\lambda_i/2} \prod_{j < k} |\lambda_i - \lambda_k|. \quad (2.1)$$

(ADMIS)

Mesure empirique. Le théorème de Marchenko-Pastur (montré par ces deux auteurs en 1967 dans l'article [5]) donne le comportement asymptotique de la mesure empirique des valeurs propres :

Théorème 2.1 (Loi de Marchenko-Pastur). *Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres de $C := \frac{1}{p} X^t X$ et $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}$. Supposons que $n \rightarrow \infty$ et $p \rightarrow \infty$ avec $n/p \rightarrow \gamma \in (0, 1]$. Alors la mesure de probabilité μ_n converge presque-sûrement pour la convergence de la topologie faible² vers la mesure de probabilité μ où μ admet pour densité*

$$\mu(dx) := \frac{\sqrt{(\gamma_+ - x)(x - \gamma_-)}}{2\pi\gamma x} 1_{x \in [\gamma_-, \gamma_+]} dx$$

avec $\gamma_{\pm} = (1 \pm \sqrt{\gamma})^2$.

Cette densité a différents comportements suivant la valeur de γ . En particulier, si $\gamma = 1$, alors $\gamma_- = 0$ et le bord gauche du spectre vaut 0 et la densité s'écrit :

$$\mu(dx) := \frac{1}{2\pi} \frac{\sqrt{4 - x^2}}{x} 1_{x \in [0,4]} dx$$

Dans ce cas, le bord gauche a une asymptote en 0. Lorsque $\gamma < 1$, le bord gauche s'annule en racine carrée comme le bord droit.

Notons que plus γ est petit i.e. p grand par rapport à n , plus la loi limite est resserrée autour de 1.

Ce théorème nous dit que dans le régime asymptotique où n et p sont du même ordre de grandeur, il est normal d'avoir des valeurs propres assez grandes (autour de γ_+ qui est strictement plus grand que 1 et qui vaut même 4 dans le cas $n/p \rightarrow 1!$). Cela contraste avec le cas où $p \gg n$, où C converge vers la matrice identité.

2. $\mu_n \rightarrow \mu$ pour la convergence de la topologie faible ssi pour toute fonction continue bornée f , on a $\int f d\mu_n \rightarrow \int f d\mu$.

Deuxième partie 2. Matrices aléatoires

2.4. Introduction rapide aux principales grandes questions des matrices aléatoires. Les matrices aléatoires X sont des matrices entrées X_{ij} aléatoires. Lorsque la matrice X est carrée, de dimension $n \times n$ et diagonalisable, on peut étudier ses valeurs propres et ses vecteurs propres (qui sont maintenant des variables aléatoires).

Loi jointe des valeurs propres. Dans certains cas particuliers où la matrice aléatoire X a de bonnes propriétés d'indépendance ou de symétrie, il est possible de trouver une expression pour la loi jointe des valeurs propres.

Dans les cas les plus célèbres, les valeurs propres vivent dans un potentiel confinant et elles se repoussent, c'est à dire qu'elles préfèrent être éloignées les unes des autres. Cette propriété remarquable a été constatée par Eugène Wigner dans les années 1950 qui a suggéré les matrices aléatoires comme un bon modèle pour étudier les atomes lourds (physique nucléaire).

Exemple 1 : Modèle GOE. Matrice symétrique $A = (X + {}^tX)/\sqrt{2}$ telle que $A_{i,j} \sim \mathcal{N}(0, 1)$ si $i > j$ et $A_{i,j} \sim \mathcal{N}(0, 2)$ et toutes les entrées sont indépendantes modulo la symétrie.

La loi jointe des valeurs propres vaut

$$P_{n,\beta}(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_n} \prod_{i < j} |\lambda_i - \lambda_j| \exp\left(-\sum_{i=1}^n \lambda_i^2/4\right)$$

Exemple 2 : Modèle Wishart. Toutes les entrées de la matrice X sont i.i.d de variance 1. Supposons de plus qu'elles sont de loi normale $\mathcal{N}(0, 1)$ (les individus sont indépendants). C'est le modèle dit de Wishart.

La densité de la loi jointe des valeurs propres de la matrice de corrélation empirique $C = \frac{1}{p} X {}^tX$ vaut :

$$f(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_{n,p}} \prod_{i=1}^n \lambda_i^{(p-n+1)/2-1} e^{-\lambda_i/2} \prod_{j < k} |\lambda_i - \lambda_k|. \quad (2.2)$$

Comportement global des valeurs propres. On s'intéresse souvent tout d'abord au **comportement global** des valeurs propres, c'est à dire à la mesure de probabilité empirique

$$\mu_n := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k}, \quad \text{où les } (\lambda_k) \text{ sont les valeurs propres de } X.$$

dans la limite où n tend vers l'infini.

Cas très simple de v.a. i.i.d. : si $\Lambda_n = (\lambda_1, \dots, \lambda_n)$ avec λ_i i.i.d de loi $\mathcal{N}(0, 1)$, on a pour la densité empirique

$$\frac{1}{n} \sum \delta_{\lambda_i} \rightarrow \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

Dans l'exemple GOE, cette mesure correctement rescalée converge presque-sûrement vers la loi du demi-cercle de Wigner.

$$\frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k/\sqrt{n}} \rightarrow \frac{1}{2\pi} \sqrt{4-x^2} \mathbf{1}_{x \in [-2,2]}$$

où la convergence a lieu pour la topologie faible.

Pour le modèle de Wishart vu ci-dessus, la mesure empirique μ_n des valeurs propres de $C = (1/p)X^t X$ converge vers la loi de Marchenko-Pastur de paramètre γ .

Comportement local des valeurs propres. Une autre information capitale concerne le **comportement local** (d'un nombre $O(1)$) des valeurs propres, par exemple des valeurs propres extrêmes (les plus grandes et les plus petites). Vers quoi converge la plus grande valeur propre (éventuellement renormalisée) lorsque $n \rightarrow \infty$? Si elle admet une limite finie, quelles sont ses fluctuations autour de cette limite? Plus précisément, on essaye de comprendre la limite microscopique des valeurs propres. Si $\Lambda_n = (\lambda_1, \dots, \lambda_n)$, on veut choisir les bonnes vitesses β_n et α_n telles que

$$\beta_n(\Lambda_n - \alpha_n) \rightarrow \text{Processus ponctuel de points (espacés de } O(1)).$$

Comment devons-nous recaler pour obtenir une limite ponctuelle lorsque $n \rightarrow \infty$?

Cas de Λ_n constitué de n v.a. i.i.d. : Si l'on choisit $\alpha_n = 0$ (on s'intéresse aux particules autour de 0), alors on a environ $O(n)$ points concentrés dans un voisinage $[-\varepsilon, \varepsilon]$ avec $\varepsilon > 0$ fixé ce qui indique qu'il faut multiplier par n pour avoir une limite non triviale.

Notons $N_n[a, b] = \#\{\text{points de } n\Lambda_n \text{ dans } [a, b]\}$

$$N_n[a, b] = \sum_{i=1}^n \mathbf{1}_{\lambda_i \in [a/n, b/n]}.$$

On a une somme de n v.a. de Bernoulli i.i.d de paramètre

$$p_n = \frac{1}{\sqrt{2\pi}} \int_{a/n}^{b/n} e^{-x^2/2} dx \sim \frac{1}{\sqrt{2\pi}} \frac{b-a}{n}.$$

Ainsi $N_n[a, b] \rightarrow \text{Poisson}((b-a)/\sqrt{2\pi})$.

On peut facilement calculer la probabilité ainsi que son asymptotique

$$\mathbb{P}[N_n[a_1, b_1] = n_1, \dots, N_n[a_k, b_k] = n_k]$$

où les intervalles $[a_i, b_i]$ sont disjoints.

On en déduit que $n \Lambda_n$ converge vers un processus de Poisson d'intensité $1/\sqrt{2\pi}$.

Si l'on choisissait $\alpha_n = c$ on trouverait $n(\Lambda_n - c)$ converge vers un processus de Poisson d'intensité $e^{-c^2/2}/\sqrt{2\pi}$.

Dans le cas des matrices aléatoires, si l'on prend le modèle GOE/ou le modèle de Wishart, la convergence de la mesure empirique nous dit environ $n\rho(c)$ points de l'ensemble Λ_n autour de c point dans le support de la mesure limite où $\Lambda_n := \{\lambda_k/\sqrt{n}, k \in \mathbb{N}\}$ dans le cas GOE et $\Lambda_n := \{\lambda_k, k \in \mathbb{N}\}$ dans le cas Wishart et $\rho(c)$ est la densité de la mesure limite au point c .

Ainsi, $\rho(c)n(\Lambda_n/\sqrt{n} - c)$ a environ $O(1)$ points dans un intervalle $[a, b]$.

Théorème 2.2.

$$2\pi\sigma(c)n\left(\frac{\Lambda_n}{\sqrt{n}} - c\right) \rightarrow \text{Sine}_1$$

où Sine_1 est un processus de points avec répulsion.

Pour les vecteurs propres, une première information capitale concerne son support : le vecteur propre répartit-il sa masse sur tout le domaine (délocalisation) ou préfère-t-il se concentrer sur un point particulier (localisation) ?

Plus grandes valeurs propres. Que peut-on maintenant dire sur les plus grandes valeurs propres ? Dans le cas où C est une matrice de Wishart, on peut déduire facilement du théorème de Marcenko-Pastur que pour tout $\varepsilon > 0$,

$$\mathbb{P}[\lambda_1 \geq \gamma_+ - \varepsilon] \rightarrow 1.$$

et donc que $\limsup \lambda_1$ est p.s. plus grande que γ_+ . En effet, choisissons une fonction f positive avec support compact inclu dans $[\gamma_+ - \varepsilon, +\infty)$ telle que $\int f d\mu = 1$. Alors

$$\begin{aligned} \mathbb{P}[\lambda_1 \leq \gamma_+ - \varepsilon] &\leq \mathbb{P}\left[\int f d\mu_n = 0\right] \\ &\leq \mathbb{P}\left[\left|\int f d\mu_n - \int f d\mu\right| \geq 1\right] \end{aligned}$$

qui tend vers 0 d'après le théorème de Marchenko-Pastur.

Mais ce théorème ne nous dit pas si la plus grande valeur propre λ converge vraiment vers γ_+ ou non. C'est facile de comprendre pourquoi : l'événement $\lambda_1 \leq \gamma_+ - \varepsilon$ implique que toutes les valeurs propres sont plus petites que $\gamma_+ - \varepsilon$, ce qui est une information globale et qui se déduit de la convergence de la mesure empirique. En revanche, un événement comme $\lambda_1 \geq \gamma_+ + \varepsilon$ implique seulement qu'une seule valeur propre dépasse la valeur $\gamma_+ + \varepsilon$ et donc est une information locale. Le théorème suivant montre qu'il n'y a pas de surprise :

Théorème 2.3 (Convergence de la plus grande valeur propre). *La plus grande valeur propre λ_1 converge p.s. vers γ_+ .*

Démonstration. Nous montrerons ce résultat plus tard en utilisant les matrices tridiagonales. □

La loi des fluctuations de λ_1 autour de γ_+ est aussi connue. Comment deviner son ordre de grandeur ?

Pour une matrice de Wishart de grande taille n , le nombre de valeurs propres dans un intervalle I_n est bien approximé par $n \times \mu[I_n]$ donc pour $I_n = [\gamma_+ - \delta_n, \gamma_+ + \delta_n]$,

$$n\mu[I_n] = n \int_0^{\delta_n} \sqrt{y} dy = n \delta_n^{3/2}$$

est d'ordre de grandeur 1 si $\delta_n = O(1/n^{2/3})$.

Le résultat suivant confirme cette prédiction :

Théorème 2.4 (Fluctuations de la première valeur propre). *La variable aléatoire*

$$\gamma^{-1/6}(1 + \sqrt{\gamma})^{-4/3} n^{2/3}(\lambda_1 - \gamma_+)$$

converge en loi vers une variable aléatoire appelée loi de Tracy Widom TW_1 quand $n \rightarrow \infty$, $p \rightarrow \infty$ avec $n/p \rightarrow \gamma$

Il est possible d’obtenir un résultat plus complet sur la loi jointe des plus grandes valeurs propres.

En reprenant l’approche processus ponctuel, c’est à dire en zoomant notre processus de points autour de son maximum, pour le cas où Λ_n est constitué de n variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$, il faut choisir $\alpha_n = 2\sqrt{\ln n} + \frac{1}{8} \frac{\ln(\pi \ln n)}{\sqrt{\ln n}}$ et $\beta_n = \sqrt{\ln n}$ et l’on obtient :

$-\beta_n(\Lambda_n - \alpha_n)$ converge vers un processus de Poisson inhomogène d’intensité $e^{-x} dx$.

Pour les valeurs propres du modèle de Wishart, on a le résultat suivant :

Théorème 2.5 (Processus de Airy). *Soit Λ_n l’ensemble des valeurs propres du modèle de Wishart, on obtient :*

$$\gamma^{-1/6}(1 + \sqrt{\gamma})^{-4/3} n^{2/3}(\Lambda_n - \gamma_+) \rightarrow \text{Airy}_1$$

2.5. Transition de Baik, Ben-Arous, Pécché. L’article [1] de Baik-Ben Arous-Pécché étudie une perturbation de rang 1 de la matrice de Wishart : que se passe-t-il si la matrice de covariance Σ du vecteur Y_1 n’est plus l’identité mais une matrice du type :

$$\Sigma = \begin{pmatrix} \ell_1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

(Les individus sont indépendants entre eux, mais on a maintenant un individu avec une variance exceptionnelle!).

La réponse dépend de la valeur de ℓ_1 .

Théorème 2.6. *Soit λ_1 la plus grande valeur propre de la matrice de covariance construite à partir de p vecteurs i.i.d. Gaussiens centrés de taille n et de matrice de covariance Σ i.e. de la matrice*

$$C = \frac{1}{p}(\sqrt{\Sigma}W)^t(\sqrt{\Sigma}W).$$

où W est une matrice de Wishart (entrées i.i.d. de loi $\mathcal{N}(0, 1)$).

- Si $\ell_1 < 1 + \sqrt{\gamma}$ alors $\lambda_1 \rightarrow \gamma_+$ avec les mêmes fluctuations que dans le cas Wishart.
- Si $\ell_1 = 1 + \sqrt{\gamma}$ alors $\lambda_1 \rightarrow \gamma_+$ avec des fluctuations d’ordre $n^{2/3}$ mais une limite différente de celle de Wishart (une loi de Tracy-Widom “déformée”).
- Si $\ell_1 > 1 + \sqrt{\gamma}$ alors $\lambda_1 \rightarrow \ell_1 + \frac{\ell_1 \gamma}{\ell_1 - 1}$ avec des fluctuations Gaussiennes d’ordre \sqrt{n} .

Intuitivement, les corrélations de la population apparaissent dans l’asymptotique de la plus grande valeur propre uniquement si elle est suffisamment grande (si elle dépasse la valeur critique $1 + \sqrt{\gamma}$). Sinon, elle disparaît devant les “fausses corrélations” qui proviennent du bruit.

Notons également que $\lambda_1 > \ell_1$ lorsque $\ell_1 > 1 + \sqrt{\gamma}$. Cela provient du fait que cette valeur propre est repoussée par les autres valeurs propres provenant de la loi de Marchenko-Pastur.

Troisième partie 3. Matrices et graphes aléatoires

Nous allons voir dans ce cours une approche géométrique introduite récemment pour étudier les matrices aléatoires. Elle nous permettra en particulier de montrer le résultat de Marchenko Pastur vu lors du dernier cours.

3. TRIGONALISATION ET MESURE SPECTRALE

Prenons une matrice symétrique A (i.e. ${}^tA = A$) de taille n . Rappelons que A est diagonalisable en base orthonormée : $A = OD{}^tO$ où O orthogonale (i.e. ${}^tOO = I_n$) qui s'écrit $O = (\varphi_1, \dots, \varphi_n)$, avec $\varphi_1, \dots, \varphi_n$ vecteurs propres de A . On suppose également dans toute cette section que le vecteur e_1 est cyclique, c'est à dire que $(e_1, Ae_1, A^2e_1, \dots, A^{n-1}e_1)$ est libre (lorsque A est choisi aléatoirement selon une loi pas fantaisiste, c'est souvent presque sûrement le cas).

Définition 3.1 (Mesure spectrale). *La mesure spectrale d'une matrice symétrique A est donnée par :*

$$\sigma_A := \sum_{i=1}^n \delta_{\lambda_i} \varphi_i(1)^2,$$

où λ_i et φ_i sont les valeurs propres et vecteurs propres de A .

Proposition 3.2 (Moments de la mesure spectrale). *Les moments de la mesure spectrale valent :*

$$\int x^k d\sigma_A(x) = (A^k)_{11}.$$

Démonstration. L'entrée 1, 1 de la matrice A^k s'écrit ${}^t e_1 A^k e_1$. Par conséquent,

$$(A^k)_{11} = {}^t e_1 O \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} {}^t O e_1$$

et

$${}^t O e_1 = \begin{pmatrix} \varphi_1(1) \\ \vdots \\ \varphi_n(1) \end{pmatrix}$$

ce qui implique le résultat. □

Un corollaire immédiat de cette proposition est que pour tout polynôme P

$$\int P(x) d\sigma_A(x) = (P(A))_{11} = {}^t e_1 P(A) e_1 .$$

Proposition 3.3. *Soit A symétrique, e_1 est cyclique ssi $|Supp(\sigma_A)| = n$*

Démonstration. Si $|\text{Supp}(\sigma_A)| < n$ alors il existe un polynôme P de degré strictement inférieur à n qui s'annule sur le support de σ_A . Or $\langle P(A)e_1, P(A)e_1 \rangle = \langle P^2(A)e_1, e_1 \rangle = \int P^2(x)d\sigma_A(x) = 0$ ce qui implique $P(A)e_1 = 0$ donc e_1 n'est pas cyclique.

Si e_1 n'est pas cyclique, alors il existe un polynôme P de degré strictement inférieur à n tel que $P(A)e_1 = 0$ donc $\int P^2(x)d\sigma_A(x) = \langle P(A)e_1, P(A)e_1 \rangle = 0$ donc la mesure σ est supportée sur moins de n points. \square

Ainsi la mesure spectrale encode les valeurs propres de A et une information partielle sur les vecteurs propres (quel poids ils mettent sur la première coordonnée). C'est une mesure de probabilités car $\sum_i \varphi_i(1)^2 = 1$ (la matrice $(\varphi_1, \dots, \varphi_n)$ est orthogonale).

Deux matrices symétriques A et B sont dites semblables s'il existe une matrice orthogonale O telle que $A = OB^tO$. C'est le cas lorsqu'elles partagent les mêmes valeurs propres (avec multiplicité) i.e. qu'elles possèdent la même mesure empirique des valeurs propres. Chaque classe de cette relation d'équivalence a un représentant simple : une matrice diagonale.

On va introduire la notion analogue pour la mesure spectrale.

Proposition 3.4. *Soient A et B deux matrices symétriques avec e_1 cyclique. On a $\sigma_A = \sigma_B$ si et seulement si il existe une matrice orthogonale O telle que $B = {}^tOAO$ où O est une matrice orthogonale de la forme $\begin{pmatrix} 1 & 0 \\ 0 & \tilde{O} \end{pmatrix}$. On dit dans ce cas que les matrices A et B sont équivalentes.*

Démonstration. Si $\sigma_A = \sigma_B$, alors A et B partagent les mêmes valeurs propres et $A = O_1D^tO_1$ et $B = O_2D^tO_2$. Donc $A = (O_1{}^tO_2)B^t(O_1{}^tO_2)$. De plus ${}^tO_1e_1 = {}^tO_2e_1$ donc $O_1{}^tO_2e_1 = e_1$ ce qui implique le résultat. Réciproquement si $B = {}^tOAO$ avec O de la bonne forme, alors A et B ont les mêmes valeurs propres. De plus si $A = O_1D^tO_1$ et $B = O_2D^tO_2$, on a $O_1 = OO_2$ et donc $(O_1)_{1k} = (O_2)_{1k}$, ainsi la première coordonnée de leurs vecteurs propres coïncide. \square

Définition 3.5 (Matrice de Jacobi). *Une matrice de Jacobi J est une matrice tridiagonale symétrique dont les entrées hors diagonales sont strictement positives.*

Proposition 3.6. *Pour toute matrice symétrique A , il existe une unique matrice de Jacobi telle que $\sigma_J = \sigma_A$.*

Ainsi, les matrices de Jacobi sont des représentants simples de la classe d'équivalence de la mesure spectrale.

Démonstration. Existence : Prenons $A = \begin{pmatrix} a & {}^tb \\ b & C \end{pmatrix}$ où $a \in \mathbb{R}, b \in \mathbb{R}^{n-1}$ et composons A par O de la forme $O = \begin{pmatrix} 1 & 0 \\ 0 & O_1 \end{pmatrix}$ (transformation de Householder). On a

$$OA^tO = \begin{pmatrix} a & {}^t(O_1b) \\ O_1b & O_1C^tO_1 \end{pmatrix}.$$

Choisissons O_1 tel que $O_1 b = |b|e_1$ alors

$$OA^tO = \begin{pmatrix} a & |b|0 & \cdots & 0 \\ |b| & & & \\ 0 & & & \\ \vdots & O_1C^tO_1 & & \\ 0 & & & \end{pmatrix}.$$

On peut itérer la procédure pour obtenir une matrice Jacobienne.

Unicité : Si J_1 et J_2 sont deux matrices Jacobiennes telles que $\sigma_{J_1} = \sigma_{J_2}$, on a $(J_1^k)_{11} = (J_2^k)_{11}$ pour tout k . En exprimant les puissances à l'aide des entrées de la matrice, il est facile de voir par récurrence que $J_1 = J_2$. □

4. CONVERGENCE DE GRAPHE

Soit $G = (V, E, w)$ un graphe localement fini avec poids $w : E \rightarrow \mathbb{R}$. Définissons l'opérateur d'adjacence A_G qui agit sur l'espace $L^2(G)$ des fonctions réelles vérifiant $\sum_{v \in V} f^2(v) < \infty$ de la manière suivante :

$$A_G(f)(v) = \sum_{e / e=(v,v') \in E} w(e)f(v'),$$

où l'on somme sur les arêtes e de la forme (v, v') et où $w(e)$ est le poids de l'arête e .

On suppose dans tout ce document que les poids sont bornés uniformément par M et que le nombre de voisins d'un sommet est lui aussi borné uniformément par D . Sous ces conditions, l'opérateur d'adjacence est borné.

Lorsque le nombre de sommets est fini de cardinal n , l'opérateur d'adjacence est représenté par une matrice d'adjacence notée A_G elle aussi. La matrice A_G est une matrice symétrique de taille $n \times n$ et la théorie spectrale nous dit qu'il existe une base orthonormale de vecteurs propres $\varphi_1, \dots, \varphi_n$ associées aux valeurs propres réelles $\lambda_1, \dots, \lambda_n$.

La mesure empirique des valeurs propres μ_G vaut $\mu_G := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}$.

On peut calculer facilement le k ième moment de μ_G comme :

$$\begin{aligned} \int x^k d\mu_G(x) &= \frac{1}{n} \text{Tr}(A_G^k) \\ &= \frac{1}{n} \sum_{v \in G} m_{G,k,v}, \end{aligned}$$

où $m_{G,k,v} = \sum_{e_1, \dots, e_k} \prod_i w(e_i)$ où (e_1, \dots, e_k) est un chemin dans G de v à v , en notant que $(A^k)_{jj}$ est la somme du produit des poids des chemins de longueur k qui commencent et reviennent en j .

Définition 4.1. Soit (G, o) un graphe enraciné fini et A_G sa matrice d'adjacence où l'entrée 1 correspond à la racine o . La mesure spectrale $\sigma_{(G,o)}$ de (G, o) est la mesure spectrale de sa matrice d'adjacence A_G .

On peut aussi calculer facilement le k -ième moment de la mesure spectrale de (G, o) . Avec la proposition 3.2, on obtient

$$\int x^k d\sigma_{(G,o)}(x) = (A_G^k)_{11} = m_{G,k,o}$$

Lorsque G est infini, les notions ci-dessus n'ont pas de sens. En particulier, il peut ne pas y avoir de valeur propre ou de vecteur propre de l'opérateur A_G . Cependant, on peut introduire la *mesure spectrale* de (G, o) par rapport à une racine $o \in V$. Une façon facile de définir cette mesure spectrale est de dire que son k ième moment vaut $m_{G,k,o}$. Il suffit ensuite de montrer qu'il existe une unique solution de ce problème (ADMIS dans ce cours, valide sous les hypothèses sur le graphe écrites ci-dessus).

En général, la mesure spectrale de (G, o) dépend du choix de la racine o . Cependant, lorsque G est sommet-transitif (i.e. quand tous les sommets jouent le même rôle à l'intérieur du graphe, par exemple lorsque G est un n -cycle ou $G = \mathbb{Z}^d$), cette mesure est indépendante de o et est appelée simplement mesure spectrale de G . Dans ce cas, cette mesure spectrale vaut μ_G la mesure empirique des valeurs propres.

Définition 4.2. *On dit que la suite de graphes enracinés (G_n, o_n) converge vers le graphe enraciné (G, o) si pour tout k , la boule de rayon k autour de la racine o_n du graphe G_n se stabilise en la boule de rayon k du graphe limite (G, o) et si les suites des poids de cette boule de taille finie convergent vers celles du graphe limite.*

Exemples :

- Le cycle de taille n (avec poids 1) converge vers \mathbb{Z} .
- la grille de taille n de \mathbb{Z}^2 (avec poids 1) converge vers \mathbb{Z}^2 .

Proposition 4.3 (Continuité de la mesure spectrale). *Si (G_n, o_n) converge vers (G, o) , alors la mesure spectrale σ_n de (G_n, o_n) converge faiblement vers σ , la mesure spectrale de (G, o) (c'est à dire que pour tout f continue bornée, on a $\int f d\sigma_n \rightarrow \int f d\sigma$).*

Remarque 4.4. *Il faut bien sûr préciser la topologie avec laquelle on travaille. On peut munir l'ensemble des graphes enracinés \mathcal{G}_o d'une distance*

$$d((G, o), (G', o')) := \sum_{r=1}^{\infty} 2^{-r} \mathbf{1}_{\mathcal{B}_r(G,o) \neq \mathcal{B}_r(G',o')}.$$

où $\mathcal{B}_r(G, o)$ désigne la boule de rayon r autour de o de G .

Démonstration. Par définition, pour tout k fixé, le moment d'ordre k de la mesure spectrale de (G_n, o_n) converge vers celui de (G, o) . Or la convergence des moments implique la convergence faible des mesures lorsque la mesure limite est déterminée par ses moments (résultat classique admis ici, voir par exemple le théorème 30.2 de [2]), ce qui est le cas ici avec nos hypothèses. \square

Exemple (calcul de la mesure spectrale de \mathbb{Z}) : On utilise la convergence du cycle de taille n vers \mathbb{Z} pour calculer la mesure spectrale de \mathbb{Z} .

Commençons par calculer la mesure spectrale du n cycle noté G_n . Sa matrice d'adjacence A_G vaut $A = T + {}^tT$ où

$$T = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ 1 & & & & 0 \end{pmatrix}$$

Le spectre de T est donné par $\{e^{2\pi ik/n}, k \in \{0, \dots, n-1\}\}$. On en déduit que le spectre de A est donné par $\{e^{2\pi ik/n} + e^{-2\pi ik/n} = 2 \operatorname{Re}(e^{2\pi ik/n}), k \in \{0, \dots, n-1\}\}$ qui correspondent à la projection sur l'axe réel de points espacés uniformément sur un cercle de rayon 2. Lorsque $n \rightarrow \infty$, le spectre deviendra la projection de la loi uniforme sur le cercle de rayon 2 sur la droite réelle, qui correspond à la loi de l'arcsinus

$$\sigma_{\mathbb{Z}}(dx) = \frac{1}{2\pi\sqrt{4-x^2}} 1_{x \in [-2,2]} dx$$

Souvent, on a à notre disposition un graphe G non enraciné. Comment généraliser le type de convergence vu ci-dessus à ces graphes ? C'est l'objet de la définition suivante :

Définition 4.5 (Convergence de Benjamini-Schramm). *Supposons que l'on ait une suite de graphes G_n finis (non enracinés). On dit que G_n converge dans le sens de Benjamini-Schramm (convergence BS) si (G_n, o_n) converge en loi vers (G, o) où o_n est une racine choisie uniformément parmi les sommets de G_n . C'est équivalent à dire que la probabilité que la boule de taille k autour de o_n de G_n vaille \mathcal{B}_k converge vers la probabilité que la boule de taille k autour de o de G vaille \mathcal{B}_k .*

Notons bien que la suite des graphes G_n est déterministe mais la limite (G, o) est aléatoire ! L'aléa de (G, o) reflète simplement l'hétérogénéité spatiale des graphes G_n .

Heuristiquement, cette convergence demande la convergence des statistiques locales du graphe, vues d'un point choisi uniformément.

Exemples de convergence BS :

- Un cycle de taille n converge vers le graphe \mathbb{Z} .
- Une chaîne de longueur n converge vers le graphe \mathbb{Z} .
- Une boîte de taille n dans \mathbb{Z}^d converge vers \mathbb{Z}^d .

(Pour les deux derniers exemples, la probabilité d'être proche du bord tend vers 0 donc on ne voit pas les effets de bords).

Dernier exemple : convergence de l'arbre binaire de taille n .

Cette convergence est particulièrement utile car elle implique la convergence de la mesure empirique des valeurs propres de G_n !

En effet grâce au théorème de convergence dominée, on a $\mathbb{E}[\sigma_{G_n, o_n}] \rightarrow \mathbb{E}[\sigma_{G, o}]$. De plus, on a $\mu_n = \mathbb{E}[\sigma_{G_n, o_n}]$ grâce à la proposition suivante :

Proposition 4.6. *Supposons que l'on ait un graphe fini G et que l'on choisisse un sommet aléatoirement uniformément parmi les sommets de G . Soit $\sigma_{(G, o)}$ sa mesure spectrale (qui est aléatoire car (G, o) est aléatoire). Alors l'espérance de $\sigma_{(G, o)}$ vaut la mesure empirique des valeurs propres du graphe d'adjacence de G .*

Démonstration. Notons n le cardinal de G . Rappelons que l'on a

$$\sigma_{(G,o)} = \sum_{i=1}^n \delta_{\lambda_i} \varphi_i(U)^2.$$

où U est le sommet choisi uniformément parmi les sommets de G . Or $\sum_x \varphi_i(x)^2 = 1$ car les vecteurs propres sont de norme 1. On en déduit

$$\mathbb{E}[\sigma_{(G,o)}] = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}.$$

□

Quatrième partie 4. Application aux matrices de Wishart : Loi de Marchenko Pastur et transition de Baik, Ben-Arous, Piché

Rappelons qu'une matrice de Wishart X de dimension $n \times p$ est une matrice où toutes les entrées sont i.i.d de loi $\mathcal{N}(0, 1)$.

Théorème 4.7. *Soit*

$$U = \frac{1}{\sqrt{p}} \begin{pmatrix} \chi_p & \tilde{\chi}_{n-1} & & & & \\ & \chi_{p-1} & \tilde{\chi}_{n-2} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{p-(n-2)} & & \\ & & & & \tilde{\chi}_1 & \\ & & & & & \chi_{p-(n-1)} \end{pmatrix}$$

avec χ et $\tilde{\chi}$ variables aléatoires indépendantes de loi χ de paramètre donné par l'indice. Alors $U^t U$ est équivalente à $C = \frac{1}{p} X^t X$ où X est une matrice de Wishart de dimension $n \times p$. En particulier ses valeurs propres admettent pour loi jointe (2.2).

Démonstration. La première étape de la preuve consiste à transformer X en matrice bi-sous-diagonale avec une opération $L = OXO'$. où $O \in \mathcal{O}(n)$ et $O' \in \mathcal{O}(p)$ sont orthogonales et de plus $Oe_1 = e_1$. Cela donne ainsi une matrice O et une matrice L telle que $X^t X = L^t L$ avec L bi-sous diagonale. Il suffit ensuite de remarquer que $U^t U$ où $U = {}^t L$ possède les mêmes valeurs propres que $L^t L$ pour conclure.

Pour la première étape :

- (1) On choisit d'abord O'_1 telle que si r_1 est la première ligne de X , ${}^t r_1 O'_1 = |r_1| e_1$ où $|r_1|$ est distribué comme χ_p . La distribution des autres lignes n'est pas changée par la multiplication par O'_1 comme celle ci est indépendante de ces lignes (O'_1 ne dépend que de r_1).
- (2) On multiplie ensuite à gauche par $O_1 = \begin{pmatrix} 1 & \\ & \tilde{O}_1 \end{pmatrix}$ qui laisse la première ligne invariante et en choisissant $\tilde{O}_1 e_1 = |c_1| e_1$ avec ${}^t (|r_1|, c_1)$ qui est la première colonne de

XO'_1 . Ainsi

$$O_1 X O'_1 = \left(\begin{array}{c|c} \begin{matrix} \chi_p \\ \tilde{\chi}_{n-1} \end{matrix} & \begin{matrix} 0 \\ \hline \end{matrix} \\ \hline \begin{matrix} 0 \\ \end{matrix} & X_1 \end{array} \right).$$

(3) On peut itérer pour continuer la procédure. □

Notons a_i et b_i les éléments diagonaux et sur-diagonaux de la matrice $\sqrt{p}U$. La matrice $U^t U$ a pour entrées :

$$\begin{aligned} (U^t U)_{i,i} &= (a_i^2 + b_i^2)/p = (\chi_{p-i+1}^2 + \tilde{\chi}_{n-i}^2)/p, \quad \text{pour } i \geq 1 \\ (U^t U)_{i,i+1} &= (a_{i+1} b_i)/p = (\chi_{p-i} \chi_{n-i})/p, \quad \text{pour } i \geq 1. \end{aligned}$$

5. CONVERGENCE DE LA PLUS GRANDE VALEUR PROPRE

Montrons dans un premier temps que la plus grande valeur propre converge vers γ_+ (Théorème 2.3) :

Lemme 5.1. *Soit J une matrice de Jacobi*

$$J = \begin{pmatrix} a_1 & b_1 & & \\ b_1 & a_2 & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ & & b_{n-1} & a_n \end{pmatrix}$$

alors sa plus grande valeur propre $\lambda_1(J)$ vérifie :

$$\lambda_1(J) \leq \max_i (a_i + b_i + b_{i-1}).$$

où $b_0 := 0$ par convention.

Démonstration. Remarquons que J peut s'écrire

$$J = -A^t A + \text{diag}(a_i + b_i + b_{i-1})$$

avec

$$A = \begin{pmatrix} 0 & \sqrt{b_1} & & \\ -\sqrt{b_1} & \sqrt{b_2} & & \\ & -\sqrt{b_2} & \sqrt{b_3} & \\ & & \ddots & \ddots \end{pmatrix}.$$

On en déduit

$$\begin{aligned} \lambda_1(J) &\leq -\lambda_1(A^t A) + \lambda_1(\text{diag}(a_i + b_i + b_{i-1})) \\ &\leq \max_i (a_i + b_i + b_{i-1}). \end{aligned}$$

où l'on a utilisé la sous-linéarité de la plus grande valeur propre (qu'on peut voir grâce à la caractérisation $\lambda_1 := \sup_x \langle Ax, x \rangle$ pour A symétrique) pour la première ligne, et la positivité de $A^t A$ pour la deuxième ligne. \square

En appliquant ce lemme à la matrice $U^t U$ définie ci-dessus, on déduit

$$\lambda_1(U^t U) \leq \frac{1}{p} \max_{i \geq 2} ((\chi_{p-i+1}^2 + \tilde{\chi}_{n-i}^2 + \chi_{p-i} \tilde{\chi}_{n-i} + \chi_{p-i+1} \tilde{\chi}_{n-i+1}), \chi_p^2 + \tilde{\chi}_{n-1}^2 + \chi_{p-1} \tilde{\chi}_{n-1})$$

En utilisant $\chi_k^2 \leq k + C\sqrt{k}$, on trouve

$$\lambda_1(U^t U) \leq \frac{1}{p}(p + n + 2\sqrt{pn}) + O\left(\frac{1}{\sqrt{p}}\right) = \gamma_+ + O\left(\frac{1}{\sqrt{p}}\right).$$

\square

Notons que cette preuve ne nous permet pas d'avoir le bon ordre de grandeur pour la fluctuation de la plus grande valeur propre (Théorème 2.4), qui est d'ordre $1/p^{2/3}$.

6. PREUVE DE LA LOI DE MARCHENKO-PASTUR

Lemme 6.1. *Supposons $n \rightarrow \infty$, $p \rightarrow \infty$ avec $n/p \rightarrow \gamma$. Considérons j_n tel que $j_n/n \rightarrow u \in [0, 1]$. Les entrées de la matrice $U^t U$ convergent vers :*

$$\begin{aligned} (U^t U)_{j_n, j_n} &\rightarrow 1 + \gamma - 2u \\ (U^t U)_{j_n, j_n+1} &\rightarrow \sqrt{\gamma - u} \sqrt{1 - u}. \end{aligned}$$

Nous laissons la preuve de ce lemme élémentaire en exercice. Ainsi le graphe de matrice d'adjacence $U^t U$ converge au sens de BS vers le graphe G_u suivant, où $u \sim \text{Unif}[0, 1]$:

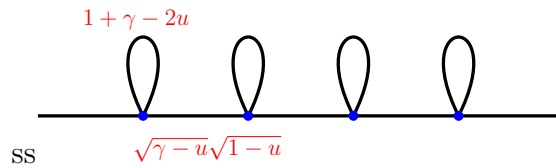


FIGURE 4. Graphe limite G_u du graphe de matrice d'adjacence $U^t U$

Il suffit donc de calculer sa mesure spectrale et de prendre sa moyenne par rapport à $u \sim \text{Unif}[0, 1]$ pour avoir la mesure limite de la mesure empirique des valeurs propres.

Notons $c_1(u) = \gamma + 1 - 2u$, $c_2(u) = \sqrt{\gamma - u} \sqrt{1 - u}$. On peut vérifier que

$$\begin{aligned} \int x^k d\sigma_{G_u} &= \sum_{\{\text{chemins allant de } 0 \text{ à } 0 \text{ de longueur } k\}} \prod (\text{poids le long du chemin}) \\ &= \sum_{i=1}^k \binom{k}{i} c_1(u)^{k-i} c_2(u)^i \#\{\text{chemins sur } \mathbb{Z} \text{ allant de } 0 \text{ à } 0 \text{ de longueur } i\} \end{aligned}$$

Or on a vu plus haut que $\int_{-2}^2 \frac{z^i}{\pi\sqrt{4-z^2}} dz$ compte le nombre de chemins de longueur i allant de l'origine à l'origine dans le graphe \mathbb{Z} (i.e. vaut $\binom{i}{i/2}$ si i pair et 0 sinon). Ce qui donne :

$$\begin{aligned} \int x^k d\sigma_{G_u} &= \sum_{i=1}^k \binom{k}{i} c_1(u)^{k-i} c_2(u)^i \int_{-2}^2 \frac{z^i}{\pi\sqrt{4-z^2}} dz \\ &= \int_{-2}^2 (c_2(u)z + c_1(u))^k \frac{1}{\pi\sqrt{4-z^2}} dz \\ &= \int x^k \mathbf{1}_{x-c_1(u) \in [-2c_2(u), 2c_2(u)]} \frac{1}{\pi\sqrt{4c_2^2(u) - (x-c_1(u))^2}} dx \end{aligned}$$

Ainsi

$$\int x^k d\sigma_{G_u} = \int x^k \mathbf{1}_{x-c_1(u) \in [-2c_2(u), 2c_2(u)]} \frac{1}{\pi\sqrt{4c_2^2(u) - (x-c_1(u))^2}} dx$$

Il nous reste à prendre la moyenne par rapport à $u \sim \text{Unif}[0, 1]$:

$$\int_0^1 \sigma_{G_u}(dx) du = \mathbf{1}_{x \in [\gamma_-, \gamma_+]} \left(\int_0^1 \frac{\mathbf{1}_{x-c_1(u) \in [-2c_2(u), 2c_2(u)]}}{\pi\sqrt{-(\gamma-1)^2 + 2x(\gamma+1) - 4xu - x^2}} du \right) dx$$

Il suffit de calculer l'intégrale pour en déduire le résultat.

7. PREUVE PARTIELLE DE LA TRANSITION DE BAIK, BEN AROUS, PÉCHÉ

On s'intéresse à la plus grande valeur propre de $\frac{1}{p}(\sqrt{\Sigma}W)^t(\sqrt{\Sigma}W)$. On tridiagonalise cette matrice afin d'obtenir une matrice symétrique $J := \frac{1}{p}B^tB$ où

$$B = \begin{pmatrix} \sqrt{\ell_1}\chi_p & \tilde{\chi}_{n-1} & & & \\ & \chi_{p-1} & \tilde{\chi}_{n-2} & & \\ & & \chi_{p-2} & \ddots & \\ & & & \ddots & \tilde{\chi}_1 \\ & & & & \chi_{p-n+1} \end{pmatrix}$$

Ainsi $J_{i,j} = (U^tU)_{i,j}$ pour tous i, j sauf $(i, j) = (1, 1)$ où $J_{1,1} = (\ell_1\chi_p^2 + \tilde{\chi}_{n-1}^2)/p$.

Si $\ell_1 \leq 1 + \sqrt{\gamma}$, il est facile d'utiliser le lemme 5.1 pour montrer que la plus grande valeur propre de J est plus petite que γ_+ ce qui montre la convergence de la plus grande valeur propre vers γ_+ .

Si $\ell_1 > 1 + \sqrt{\gamma}$,

En utilisant $\chi_k^2 \sim k + O(\sqrt{k})$ lorsque $k \gg 1$, on approxime le haut de la matrice J par la matrice déterministe

$$\begin{pmatrix} \ell_1 + \gamma & \sqrt{\gamma} & & & \\ \sqrt{\gamma} & 1 + \gamma & \sqrt{\gamma} & & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots \end{pmatrix} + O\left(\frac{1}{\sqrt{p}}\right)$$

qui est une approximation du Laplacien rescalé et recentré.

Soit $a = \sqrt{\gamma}/(\ell_1 - 1) < 1$ et $\ell_1^+ := \ell_1 + \gamma(\ell_1/(\ell_1 - 1))$ (qui est la position attendue de la première valeur propre de J). Considérons le vecteur $v_k = (1, a, a^2, \dots, a^k, 0, \dots, 0)$. On a

$$\|Jv_k - \ell_1^+ v_k\| \leq Ca^k + O(1/\sqrt{p})$$

Or on a le lemme suivant :

Lemme 7.1. *Si une matrice symétrique A qui vérifie $\|Ax - \lambda x\| \leq \varepsilon$ pour x vecteur de taille supérieure à 1, il existe une valeur propre de A à distance inférieure à ε de λ .*

On en déduit que J a une valeur propre à distance inférieure à a^k de ℓ_1^+ .

Il faut maintenant montrer que c'est bien la plus grande valeur propre.

Lemme 7.2 (Inégalité de Weyl). *Soit A une matrice symétrique et $A' = A + R$ une perturbation de cette matrice par R matrice de rang 1 positive. Alors les valeurs propres de A et A' s'entrelacent et le décalage des valeurs propres de A' est sur la droite.*

Ce lemme implique que la deuxième valeur propre λ_2 de J est inférieure à la plus grande valeur propre de $U^t U$ or on a vu que cette plus grande valeur propre converge vers γ_+ . On en déduit le résultat.

8. LIMITES DES PROCESSUS ET HEURISTIQUE DE CONVERGENCE

Pour avoir le résultat de la convergence vers un processus d'Airy, considérons le cas $\ell_1 = 1$ en premier (le cas $\ell_1 \neq 1$ est plus complexe mais les mêmes heuristiques peuvent être utilisées).

On reprend l'approximation de la matrice vue ci-dessus et on utilise

$$\chi_k \simeq \sqrt{k} + \sqrt{1/2}\mathcal{N}(0, 1)$$

$$\chi_k^2 \simeq k + \sqrt{2k}\mathcal{N}(0, 1)$$

Les matrices tridiagonales sont considérées comme des opérateurs différentiels discrets. Pour tout $\alpha \leq 1$, nous dirons qu'une suite de matrices $(A_n, n \in \mathbb{N})$ converge vers un opérateur A (agissant sur les fonctions) et noterons $A_n \xrightarrow{n^\alpha} A$ lorsque pour tout f "régulière"³ à support compact et à valeurs positives, on a

$$\|A_n f(\cdot/n^\alpha) - A f(\cdot/n^\alpha)\|_\infty \rightarrow 0$$

(pour chaque n , f est identifiée au vecteur $(f(1/n^\alpha), \dots, f(n/n^\alpha))$ et A_n agit sur ce vecteur. La notation $A f(\cdot/n^\alpha)$ représente le vecteur $(A f(1/n^\alpha), \dots, A f(n/n^\alpha))$).

Ainsi avec ces notations,

$$\begin{bmatrix} 2 & 0 & \cdots & 0 \\ 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2 \end{bmatrix} \xrightarrow{\sqrt{n}} \times \text{par } 2 \quad ; \quad \begin{bmatrix} 0 & 2 & \cdots & 0 \\ 0 & 0 & 2 & \vdots \\ \vdots & \ddots & \ddots & 2 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \xrightarrow{\sqrt{n}} \times \text{par } 2$$

³. Nous sommes volontairement flous sur cette définition car nous ne présentons que des arguments heuristiques.

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & n \end{bmatrix} \xrightarrow{\sqrt{n}} \times \text{par } x ; \quad \sqrt{n} \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \xrightarrow{\sqrt{n}} \frac{d}{dx} \\
& n \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & 1 & -2 & 1 \end{bmatrix} \xrightarrow{\sqrt{n}} \frac{d^2}{dx^2} ; \quad n^{1/4} \begin{bmatrix} g_1 & 0 & \cdots & 0 \\ 0 & g_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & g_n \end{bmatrix} \text{ " } \xrightarrow{\sqrt{n}} \text{ " } \times \text{ par } B'
\end{aligned}$$

où pour la dernière convergence, les g_k sont des variables aléatoires indépendantes de loi gaussienne centrée réduite.

A COMPLETER.

RÉFÉRENCES

- [1] Baik, Ben Arous, Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*. Ann. Proba., 2005.
- [2] Billingsley, *Probability and measure, third ed.*, Wiley Series in Probability and Mathematical Statistics, 1995.
- [3] Bloemendal, Virág, *Limits of spiked random matrices I*, Probability Theory and Related Fields, 2013.
- [4] Bouchaud, Potters, *Financial Applications of Random Matrix Theory : a short review*, arXiv.org, Quantitative Finance Papers, 2009.
- [5] Marchenko, Pastur, *Distribution of eigenvalues for some sets of random matrices*, Maths USSR SB, 1967.
- [6] Novembre et al., *Genes Mirror Geography within Europe*, Nature, 456, p. 98–101, 2008

UNIVERSITÉ PARIS-DAUPHINE, PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY, 75016 PARIS
E-mail address: dumaz@ceremade.dauphine.fr