

UNIVERSITÉ DE RENNES 1

ÉCOLE NORMALE SUPÉRIEURE DE CACHAN
ANTENNE DE BRETAGNE

RAPPORT DE STAGE

Stage de fin de 1^{ère} année de Magistère de Mathématiques
Du 7 mai au 22 juin 2012

ÉTUDE DE SPECTRES DE MATRICES SYMÉTRIQUES ALÉATOIRES À COEFFICIENTS CORRÉLÉS

Salim ROSTAM

Sous la direction de Gérard LE CAËR et de Renaud DELANNAY
Institut de Physique de Rennes
Département Milieux Divisés



Table des matières

Introduction	4
1 L'ensemble Gaussien Orthogonal	5
1.1 Caractérisation des matrices GOE	5
1.2 Répartition des valeurs propres	10
1.2.1 Premiers essais	10
1.2.2 Le théorème de Wigner	12
2 Distribution des valeurs propres : cas uniformément corrélé	13
3 Distribution des valeurs propres : corrélation par blocs	20
3.1 Nombre de blocs constant	20
3.2 Taille des blocs constante	24
3.3 Cas pathologiques	27
Conclusion	28
A Création des différentes matrices	30
A.1 Cas GOE	30
A.2 Cas de corrélations par blocs identiques	30
B Graphiques	32
B.1 Histogramme	32
B.2 Ajustement par la méthode des moindres carrés	32
B.3 Temps de calcul	33
C Densité jointe des valeurs propres des matrices GOE	33
D Une loi du χ^2	34
E Termes d'ajouts non gaussiens	35
E.1 Uniforme discret	35
E.2 Uniforme continu	39

Remerciements

Merci à M. Gradinaru de m'avoir fait connaître ce stage, merci à Renaud Delannay et à Gérard Le Caër de me l'avoir proposé, merci à Gérard Le Caër de m'avoir expliqué tant de choses, merci à David, Imen, Lucie et Mickaël de m'avoir accueilli dans le « Thésarland » de l'IPR.

Introduction

(Cette introduction est largement inspirée de [1].) L'étude des matrices aléatoires et en particulier de leurs éléments propres a véritablement débuté dans les années 1950 pour étudier la résonance de neutrons lents dans les noyaux lourds. On doit alors étudier la célèbre équation de Schrödinger $H\Psi = E\Psi$: les différents niveaux d'énergie d'un noyau sont alors donnés par les *valeurs propres* de l'opérateur hermitien H , appelé Hamiltonien. Cet opérateur étant très complexe (et accessoirement pouvant être considéré comme opérant dans un espace de Hilbert de dimension infinie), on le modélise par une matrice de *très grande taille* : les dimensions restantes sont négligées. Wigner est le premier à étudier l'analogie avec les valeurs propres de matrices aléatoires ; ainsi, il remplace en quelque sorte les valeurs propres de l'opérateur H très complexe par celles d'une matrice aléatoire de très grande taille (au moins plusieurs milliers).

Par la suite, les matrices aléatoires sont apparues dans d'autres domaines de la physique comme la caractérisation des systèmes chaotiques, l'étude de la conductivité dans les matériaux désordonnés, la gravitation quantique ou encore la théorie des cordes ; elles apparaissent également dans les mathématiques, par exemple pour rechercher des permutations qui possèdent certaines particularités ou bien pour étudier la répartition des zéros non triviaux de la fonction ζ de Riemann sur la ligne critique d'équation $\text{Re } z = \frac{1}{2}$. Plus précisément, la distribution des corrélations de paires¹ des zéros sur la droite critique est la même que celle des corrélations de paires des valeurs propres d'une matrice hermitienne aléatoire ; en l'occurrence, cette fonction est donnée par $r \mapsto 1 - \text{sinc}^2(\pi r)$. Cela conforte l'idée que les zéros de la fonction ζ sur la ligne critique seraient en fait les valeurs propres d'un certain opérateur hermitien : c'est la conjecture de Hilbert-Pólya.

Plus surprenant encore, G. Le Caër a montré en 1989 que la distribution des arbres dans les forêts scandinaves ressemblait à celle des valeurs propres d'une matrice complexe aléatoire, ce qui n'est en revanche pas le cas pour la distribution des chefs-lieux français (G. Le Caër et R. Delannay, 1993). Finalement, il est important de préciser que l'on ne comprend pas encore très bien pourquoi les matrices aléatoires sont des objets adéquats pour modéliser, simplifier ces différentes situations.

Comme le dit Mehta lui-même dans [1], les coefficients des matrices aléatoires sont pris *indépendants* pour des raisons de simplicité et non pas pour des raisons physiques, où ils ont toutes les chances d'être dépendants entre eux : Mehta parle de « mathematical idealization ». Malgré l'article d'Arnold [2], on s'est longtemps restreint aux matrices aléatoires avec des coefficients indépendants. Depuis quelques années seulement, on commence à étudier ce qui se passe lorsque l'on introduit des corrélations entre les coefficients : voir par exemple [4], [5], [6], [7] où les corrélations étudiées sont inspirées de la structure des matrices de Toeplitz (corrélations selon les diagonales).

C'est dans cette optique de l'étude des cas corrélés que ce stage se situe. Dans ce rapport, après avoir rappelé des résultats bien connus sur les matrices symétriques réelles à coefficients indépendants, on s'intéressera à des structures de corrélations particulières

1. Une « two-point correlation function » est une fonction $g(r)$ qui donne la probabilité de trouver deux points distants de r .

où la matrice de corrélation est *diagonale par blocs* avec des blocs identiques; on remarquera que le fait que les blocs soient identiques exclut le cas Toeplitz par exemple. On s'intéressera en particulier à la distribution *asymptotique* (quand la taille des matrices tend vers l'infini) à une valeur propre.

Dans la suite de ce rapport, toutes les variables et matrices considérées sont réelles.

1 L'ensemble Gaussien Orthogonal

Dans toute cette section, n désigne un entier naturel non nul et on note N l'entier $\frac{n(n+1)}{2}$.

Définition 1. Soit $(M_{i,j})_{1 \leq j \leq i \leq n}$ une famille de variable aléatoires *indépendantes* telles que :

- pour $1 \leq i \leq n$, $M_{i,i}$ suit une loi $\mathcal{N}(m, 2\sigma^2)$;
- pour $1 \leq i < j \leq n$, $M_{i,j}$ suit une loi $\mathcal{N}(0, \sigma^2)$;

où m et σ sont deux réels. On dit que la matrice aléatoire M définie par $M = (M_{i,j})_{1 \leq i, j \leq n}$ où $M_{i,j} = M_{j,i}$ pour $i > j$ fait partie de l'*Ensemble Gaussien Orthogonal* (Gaussian Orthogonal Ensemble). De plus, on dira que les coefficients $(M_{i,j})_{1 \leq i \leq j \leq n}$ sont les coefficients *sur-diagonaux* de M .

1.1 Caractérisation des matrices GOE

Théorème 2. *Les matrices GOE sont les seules matrices aléatoires symétriques² M qui vérifient les deux conditions suivantes :*

- (i) *les coefficients sur-diagonaux de M sont indépendants ;*
- (ii) *si R est une matrice orthogonale, alors M et $RM R^T$ ont la même loi.*

La démonstration va se dérouler en deux étapes : on va d'abord montrer — après quelques préparatifs — que les matrices GOE vérifient ces conditions, puis que ces conditions conduisent à des matrices GOE (cette fois pour $n = 2$ uniquement, pour des raisons de clarté).

Définition 3. Soit M une matrice symétrique de taille n . On note $\text{vech}(M)$ le vecteur de \mathbb{R}^N constitué des coefficients sur-diagonaux de M pris par diagonales où les coefficients ont été ajustés pour que la *norme euclidienne soit conservée*. Plus précisément :

$$\text{vech}(M) = (m_{1,1}, \dots, m_{n,n}, \sqrt{2}m_{1,2}, \dots, \sqrt{2}m_{n-1,n}, \dots, \sqrt{2}m_{1,n})^T.$$

Note. Par « norme euclidienne », on entend $\|x\|_2^2 = \sum_i x_i^2$ et $\|M\|_2^2 = \sum_{i,j} m_{i,j}^2$. En outre, si M est symétrique on a $\|M\|_2^2 = \text{tr}(M^2)$.

Remarque 4. L'opérateur vech réalise un isomorphisme entre $\mathcal{S}_n(\mathbb{R})$ et \mathbb{R}^N .

2. En toute rigueur, à densité dérivable.

Lemme 5. Soit U une matrice orthogonale de taille n . Alors :

$$\exists! V_U \in \mathcal{M}_N(\mathbb{R}), \forall M \in \mathcal{S}_n(\mathbb{R}), \text{vech}(UMU^T) = V_U \text{vech}(M).$$

De plus, la matrice V_U est orthogonale.

Note. La matrice UMU^T est symétrique : on peut donc bien parler de son vech .

Remarque 6. Les matrices $(UMU^T)^2$ et M^2 étant semblables, elles ont même trace donc UMU^T et M ont même norme euclidienne.

Démonstration du lemme.

(Existence.) On a :

$$\begin{aligned} (UMU^T)_{i,j} &= \sum_{k,l} u_{i,k} m_{k,l} u_{l,j}^T \\ &= \sum_{k,l} u_{i,k} u_{j,l} m_{k,l} \\ (UMU^T)_{i,j} &= \sum_k u_{i,k} u_{j,k} m_{k,k} + 2 \sum_{k<l} u_{i,k} u_{j,k} m_{k,l} \end{aligned} \quad (1)$$

qui est une combinaison linéaire des coefficients de $\text{vech}(M)$: de plus, les coefficients de la combinaison linéaire ne dépendent que de la matrice U . L'application $\text{vech}(M) \mapsto \text{vech}(UMU^T)$ est donc une application linéaire qui ne dépend que de U : en notant V_U la matrice de cette application linéaire dans la base canonique, on a bien $\text{vech}(UMU^T) = V_U \text{vech}(M)$ pour toute matrice M symétrique.

(Unicité.) Si $V \in \mathcal{M}_N(\mathbb{R})$ est telle que $\forall M \in \mathcal{S}_n(\mathbb{R}), V \text{vech}(M) = \text{vech}(UMU^T)$ alors on a $\forall M \in \mathcal{S}_n(\mathbb{R}), V \text{vech}(M) = V_U \text{vech}(M)$ et donc d'après la remarque 4, $\forall x \in \mathbb{R}^N, Vx = V_U x$ d'où $V = V_U$.

(V_U est orthogonale.) Soit $x \in \mathbb{R}^N$; toujours d'après la remarque 4, on peut écrire $x = \text{vech}(M)$ où M est une matrice symétrique de taille n . On a donc :

$$\begin{aligned} \|V_U x\|_2 &= \left\| \text{vech}(UMU^T) \right\|_2 \\ &= \left\| UMU^T \right\|_2 \quad (\text{par construction du vech}) \\ &= \|M\|_2 \quad (\text{d'après la remarque 6}) \\ &= \|\text{vech}(M)\|_2 \\ \|V_U x\|_2 &= \|x\|_2. \end{aligned}$$

Ainsi, V_U conserve la norme euclidienne donc V_U est orthogonale. □

Remarque 7. Soit U une matrice orthogonale. Alors si $y = V_U x$, on a $\text{vech}^{-1}(y) = U \text{vech}^{-1}(x) U^T$; il suffit par exemple d'expliciter les relations entre les coefficients.

On est maintenant prêt pour démontrer le théorème !

Démonstration du théorème : première étape. Soit M une matrice GOE de taille n : on va montrer que M vérifie la condition (ii). On note $M_{i,i} \sim \mathcal{N}(m, 2\sigma^2)$ et $M_{i,j} \sim \mathcal{N}(0, \sigma^2)$. Les coefficients de $\text{vech}(M)$ sont des gaussiennes indépendantes donc $\text{vech}(M)$ est un vecteur gaussien. Son vecteur espérance est $\mu = (\underbrace{m, \dots, m}_{n \text{ fois}}, 0, \dots, 0)^T$ et sa matrice de covariance est l'identité.

Soit maintenant U une matrice orthogonale ; on pose $M' = U M U^T$. Si V_U est comme dans le lemme 5, on a $\text{vech}(M') = V_U \text{vech}(M)$. Le vecteur $\text{vech}(M')$ est donc gaussien, de matrice de covariance $V_U I V_U^T = I$ car V_U est orthogonale. Pour déterminer son vecteur espérance, donné par $V_U \mu$, on va utiliser l'expression (1). Compte-tenu de la forme de μ , il suffit de déterminer pour chaque ligne de V_U la valeur de la somme des n premiers coefficients. D'après (1), cette somme s'écrit $\sum_k u_{i,k} u_{j,k}$: or, U est orthogonale donc cette somme vaut $\delta_{i,j}$. Finalement, le vecteur espérance de $\text{vech}(M')$ est $(\underbrace{1 \times m, \dots, 1 \times m}_{n \text{ fois}}, \cdot \times 0, \dots, \cdot \times 0)^T = \mu$! Les vecteurs gaussiens $\text{vech}(M')$ et $\text{vech}(M)$ ont même vecteur espérance et même matrice de covariance : ils sont donc même distribution, donc M' et M également ! \square

Démonstration du théorème : deuxième étape. Soit M une matrice aléatoire symétrique 2×2 qui vérifie les conditions (i) et (ii). On pose :

$$R_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

(note³) puis $N_\theta = R_\theta M R_\theta^T$. Puisque R_θ est orthogonale, on peut appliquer le lemme 5 : on note \mathcal{R}_θ la matrice orthogonale de taille 3×3 telle que $\text{vech}(N_\theta) = \mathcal{R}_\theta \text{vech}(M)$. On a alors :

$$\begin{aligned} \mathbb{E}[g(\text{vech}(N_\theta))] &= \mathbb{E}[g(\mathcal{R}_\theta \text{vech}(M))] \\ &= \int g(\mathcal{R}_\theta x) f_M(x) dx \\ &= \int g(y) f_M(\mathcal{R}_\theta^T y) |\det \mathcal{R}_\theta^T| dy \\ \mathbb{E}[g(\text{vech}(N_\theta))] &= \int g(y) f_M(\mathcal{R}_\theta^T y) dy \end{aligned}$$

(le $|\det|$ valant 1 car la matrice \mathcal{R}_θ^T est orthogonale.) Ainsi, N est à densité et :

$$f_{N_\theta}(y) = f_M(\mathcal{R}_\theta^T y).$$

La condition (ii) nous dit :

$$\forall y, \frac{\partial f_{N_\theta}(y)}{\partial \theta} = \frac{\partial f_M(\mathcal{R}_\theta^T y)}{\partial \theta} = 0. \quad (2)$$

3. Le cas général $n \geq 2$ consiste uniquement à considérer les matrices du type $\text{diag}(I_p, R_\theta, I_q)$.

Cette égalité « brute » étant difficile à exploiter, on va poser pour $y \in \mathbb{R}^3$:

$$x(y, \theta) = \mathcal{R}_\theta^T y.$$

Pour des raisons de lisibilité, on omettra dans la suite de préciser les arguments (y, θ) de x . D'après la condition (i) d'indépendance, on peut écrire :

$$f_M(x) = f_1(x_1)f_2(x_2)f_3\left(\frac{x_3}{\sqrt{2}}\right)$$

où f_1, f_2 sont les lois de $M_{1,1}, M_{2,2}$ et où f_3 est la loi de $M_{1,2}$. On a donc $f_{N_\theta}(y) = f_1(x_1)f_2(x_2)f_3\left(\frac{x_3}{\sqrt{2}}\right)$ et l'équation (2) se réécrit de la façon suivante :

$$\begin{aligned} \frac{\partial f_{N_\theta}(y)}{\partial \theta} &= f_1'(x_1) \frac{\partial x_1}{\partial \theta} f_2(x_2) f_3\left(\frac{x_3}{\sqrt{2}}\right) + f_1(x_1) f_2'(x_2) \frac{\partial x_2}{\partial \theta} f_3\left(\frac{x_3}{\sqrt{2}}\right) \\ &\quad + \frac{1}{\sqrt{2}} f_1(x_1) f_2(x_2) f_3'\left(\frac{x_3}{\sqrt{2}}\right) \frac{\partial x_3}{\partial \theta} = 0 \end{aligned}$$

d'où en divisant par $f_M(x)$:

$$\frac{f_1'(x_1)}{f_1(x_1)} \frac{\partial x_1}{\partial \theta} + \frac{f_2'(x_2)}{f_2(x_2)} \frac{\partial x_2}{\partial \theta} + \frac{1}{\sqrt{2}} \frac{f_3'\left(\frac{x_3}{\sqrt{2}}\right)}{f_3\left(\frac{x_3}{\sqrt{2}}\right)} \frac{\partial x_3}{\partial \theta} = 0. \quad (3)$$

Il reste maintenant à expliciter les quantités $\frac{\partial x_i}{\partial \theta}$. On a $y = \mathcal{R}_\theta x$ donc d'après la remarque 7, en notant $Y = \text{vech}^{-1}(y)$ et $X = \text{vech}^{-1}(x)$ on a $Y = R_\theta X R_\theta^T$ ce que l'on écrit :

$$X = R_\theta^T Y R_\theta. \quad (4)$$

Note. Attention au $\sqrt{2}$: on a $X = \begin{pmatrix} x_1 & \frac{x_3}{\sqrt{2}} \\ \frac{x_3}{\sqrt{2}} & x_2 \end{pmatrix}$.

En dérivant l'égalité (4) :

$$\begin{aligned} \frac{\partial X}{\partial \theta} &= \frac{\partial R_\theta^T}{\partial \theta} Y R_\theta + R_\theta^T Y \frac{\partial R_\theta}{\partial \theta} \\ &= \frac{\partial R_\theta^T}{\partial \theta} R_\theta X + X R_\theta^T \frac{\partial R_\theta}{\partial \theta} \\ \frac{\partial X}{\partial \theta} &= \left(R_\theta^T \frac{\partial R_\theta}{\partial \theta} \right)^T X + X \left(R_\theta^T \frac{\partial R_\theta}{\partial \theta} \right) \end{aligned} \quad (5)$$

Or,

$$R_\theta^T \frac{\partial R_\theta}{\partial \theta} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} -\sin \theta & \cos \theta \\ -\cos \theta & -\sin \theta \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

L'égalité (5) se réécrit donc ainsi :

$$\begin{aligned}\frac{\partial X}{\partial \theta} &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 & \frac{x_3}{\sqrt{2}} \\ \frac{x_3}{\sqrt{2}} & x_2 \end{pmatrix} + \begin{pmatrix} x_1 & \frac{x_3}{\sqrt{2}} \\ \frac{x_3}{\sqrt{2}} & x_2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{x_3}{\sqrt{2}} & -x_2 \\ x_1 & \frac{x_3}{\sqrt{2}} \end{pmatrix} + \begin{pmatrix} -\frac{x_3}{\sqrt{2}} & x_1 \\ -x_2 & \frac{x_3}{\sqrt{2}} \end{pmatrix} \\ \frac{\partial X}{\partial \theta} &= \begin{pmatrix} -\sqrt{2}x_3 & x_1 - x_2 \\ x_1 - x_2 & \sqrt{2}x_3 \end{pmatrix}\end{aligned}$$

On en déduit les relations suivantes :

$$\begin{cases} \frac{\partial x_1}{\partial \theta} = -\sqrt{2}x_3 \\ \frac{\partial x_2}{\partial \theta} = \sqrt{2}x_3 \\ \frac{\partial x_3}{\partial \theta} = \sqrt{2}(x_1 - x_2) \end{cases}$$

En les injectant dans l'égalité (3), on a :

$$-\sqrt{2} \frac{f'_1(x_1)}{f_1(x_1)} x_3 + \sqrt{2} \frac{f'_2(x_2)}{f_2(x_2)} x_3 + \frac{f'_3(\frac{x_3}{\sqrt{2}})}{f_3(\frac{x_3}{\sqrt{2}})} (x_1 - x_2) = 0.$$

En réarrangeant un peu, cela donne :

$$\frac{1}{\sqrt{2}x_3} \frac{f'_3(\frac{x_3}{\sqrt{2}})}{f_3(\frac{x_3}{\sqrt{2}})} = \frac{1}{x_1 - x_2} \left(\frac{f'_1(x_1)}{f_1(x_1)} - \frac{f'_2(x_2)}{f_2(x_2)} \right). \quad (6)$$

Il est maintenant temps de se rappeler que $x = x(y, \theta)$ et que $X(y, \theta) = R_\theta^T Y R_\theta$ où les majuscules sont les vech⁻¹ des minuscules. On a établi l'égalité (6) pour tout (y, θ) . En prenant $\theta = 0$, on a $R_\theta = I_2$ d'où $X(y, 0) = Y$ donc $x(y, 0) = y$. Finalement :

$$\forall y \in \mathbb{R}^3, \frac{1}{\sqrt{2}y_3} \frac{f'_3(\frac{y_3}{\sqrt{2}})}{f_3(\frac{y_3}{\sqrt{2}})} = \frac{1}{y_1 - y_2} \left(\frac{f'_1(y_1)}{f_1(y_1)} - \frac{f'_2(y_2)}{f_2(y_2)} \right).$$

Chaque membre de l'égalité dépend de variables différentes, ces deux membres sont donc *constants*. On obtient alors deux équations :

$$\begin{cases} \frac{1}{\sqrt{2}y_3} \frac{f'_3(\frac{y_3}{\sqrt{2}})}{f_3(\frac{y_3}{\sqrt{2}})} = -C_1 \end{cases} \quad (7a)$$

$$\begin{cases} \frac{1}{y_1 - y_2} \left(\frac{f'_1(y_1)}{f_1(y_1)} - \frac{f'_2(y_2)}{f_2(y_2)} \right) = -C_1 \end{cases} \quad (7b)$$

L'équation (7a) se réécrit $\frac{1}{\sqrt{2}} f'_3(\frac{y_3}{\sqrt{2}}) = -C_1 y_3 f_3(\frac{y_3}{\sqrt{2}})$ d'où $f_3(\frac{y_3}{\sqrt{2}}) = C_2 \exp(-C_1 \frac{y_3^2}{2})$ et finalement :

$$f_3(y_3) = C_2 \exp(-C_1 y_3^2).$$

La densité f_3 est intégrable (puisque c'est une densité), donc $C_1 \geq 0$. On reconnaît alors une distribution gaussienne :

$$M_{1,2} \sim \mathcal{N}\left(0, \frac{1}{2C_1}\right).$$

En séparant les variables dans l'équation (7b) on obtient par le même artifice que précédemment :

$$\begin{cases} \frac{f'_1(y_1)}{f_1(y_1)} + C_1 y_1 = C_3 \\ \frac{f'_2(y_2)}{f_2(y_2)} + C_1 y_2 = C_3 \end{cases}$$

et on en déduit :

$$\begin{cases} f_1(y_1) = C_4 \exp\left(C_3 y_1 - C_1 \frac{y_1^2}{2}\right) \\ f_2(y_2) = C_4 \exp\left(C_3 y_2 - C_1 \frac{y_2^2}{2}\right) \end{cases}$$

On reconnaît deux distributions gaussiennes, d'où :

$$M_{1,1} \sim M_{2,2} \sim \mathcal{N}\left(-\frac{C_3}{C_1}, \frac{1}{C_1}\right)$$

On retrouve alors la forme annoncée. □

1.2 Répartition des valeurs propres

1.2.1 Premiers essais

Le plus facile à faire est de générer des tirages d'une matrice GOE puis de tracer l'histogramme des valeurs propres ; les programmes en Scilab sont détaillés dans l'annexe A.1. La première chose que l'on peut essayer de faire est de prendre $\sigma^2 = 1$. On obtient alors par exemple la figure 1 : on peut raisonnablement penser que les valeurs propres se répartissent suivant un *demi-cercle*, que l'on a ajusté par moindres carrés⁴ ; on voit cependant que la concordance n'est pas parfaite.

Or, il se trouve que toutes les valeurs propres ont la même loi⁵. Les valeurs propres auraient donc une densité égale à :

$$x \mapsto \frac{2}{\pi a^2} \sqrt{a^2 - x^2} \mathbf{1}_{|x| \leq a}$$

4. Se reporter à l'annexe B.2 pour plus de détails.

5. Se reporter à l'annexe C pour plus de détails.

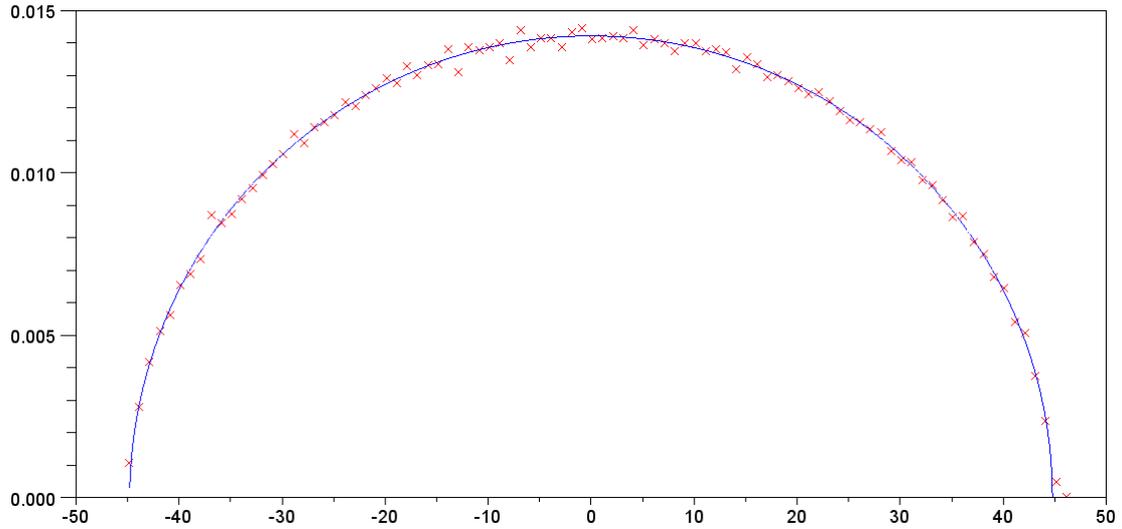


FIGURE 1 – Répartition des valeurs propres pour 100 matrices GOE de taille 500

où a est le rayon du demi-cercle. On peut calculer le moment d'ordre 2 :

$$\begin{aligned}
 \mathbb{E}[\lambda^2] &= \int_{-a}^a x^2 \frac{2}{\pi a^2} \sqrt{a^2 - x^2} dx \\
 &= \frac{2}{\pi} \int_{-a}^a \left(\frac{x}{a}\right)^2 a \sqrt{1 - \left(\frac{x}{a}\right)^2} dx \\
 &= \frac{2a^2}{\pi} \int_{-1}^1 u^2 \sqrt{1 - u^2} du \\
 &= \frac{2a^2}{\pi} \int_0^1 u \sqrt{1 - u^2} \times 2u du \\
 \mathbb{E}[\lambda^2] &= \frac{2a^2}{\pi} \int_0^1 \sqrt{t} \sqrt{1 - t} dt
 \end{aligned}$$

On remarque que cette dernière intégrale est en fait $\beta(\frac{3}{2}, \frac{3}{2})$: elle vaut donc $\frac{\pi}{8}$. On trouve finalement :

$$\mathbb{E}[\lambda^2] = \frac{a^2}{4}. \quad (8)$$

On a donc $\mathbb{E}[\text{tr}(M^2)] = \mathbb{E}[\sum \lambda^2] = \sum \mathbb{E}[\lambda^2] = \frac{na^2}{4}$. Or, on sait que :

$$\begin{aligned}
 \mathbb{E}[\text{tr}(M^2)] &= \sum_i \mathbb{E}[m_{i,i}^2] + \sum_{i \neq j} \mathbb{E}[m_{i,j}^2] \\
 &= n \times 2\sigma^2 + n(n-1)\sigma^2 \\
 \mathbb{E}[\text{tr}(M^2)] &= n(n+1)\sigma^2.
 \end{aligned}$$

On conclut que $\frac{na^2}{4} = n(n+1)\sigma^2$ i.e. $\frac{a^2}{4} = (n+1)\sigma^2$. On choisit de prendre $a = 1$:

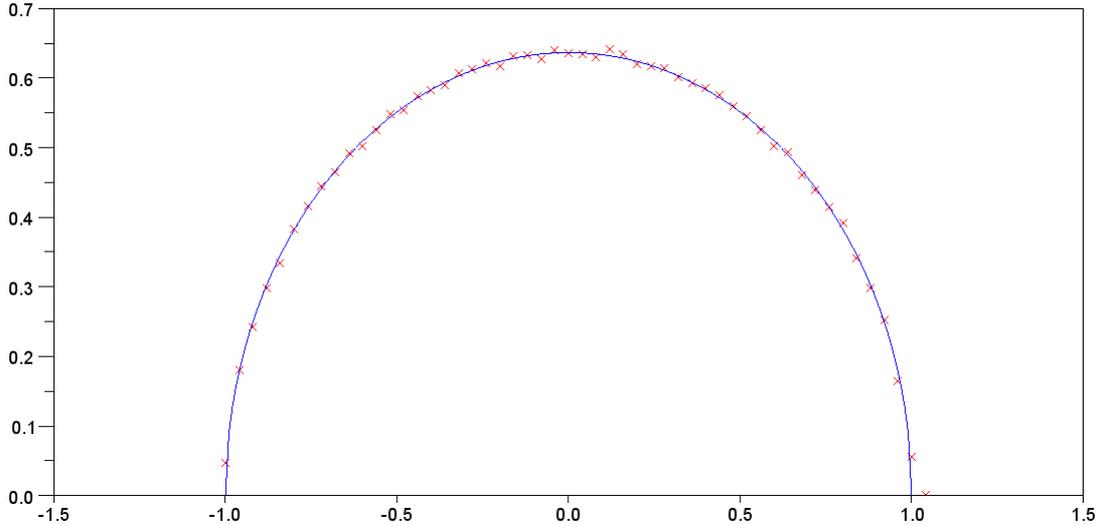


FIGURE 2 – Répartition des valeurs propres pour 100 matrices GOE de taille 500

on a donc forcément $\sigma^2 = \frac{1}{4(n+1)}$. On peut vérifier sur la figure 2 que l’histogramme coïncide presque avec le demi-cercle centré de rayon 1.

Dans toute la suite (sauf mention du contraire), on aura $\sigma^2 = \frac{1}{4(n+1)}$ et $m = 0$.

1.2.2 Le théorème de Wigner

L’observation n’est en fait ici pas trompeuse : le théorème de Wigner dit justement que, quand la taille des matrices tend vers l’infini alors les valeurs propres se répartissent sur un demi-cercle ! Plus précisément, la densité à une valeur propre est asymptotiquement le demi-cercle centré de rayon 1.

Définition 8. Soit M une matrice aléatoire de taille n . La *fonction de répartition empirique aléatoire* des valeurs propres est définie par :

$$\sigma_n(\omega, x) = \frac{1}{n} \#\{\lambda \in \text{Sp}(M(\omega)) / \lambda \leq x\}.$$

Pour un ω donné, la fonction $\sigma_n(\omega, \cdot)$ est une fonction de répartition : la loi de probabilité associée est $\mu_n(\omega, \cdot) = \frac{1}{n} \sum_{\lambda \in \text{Sp}(M(\omega))} \delta_\lambda$.

Théorème 9. Soit M une matrice GOE de taille n . Alors la mesure aléatoire μ_n converge étroitement en probabilité vers la densité du demi-cercle $\rho(x) = \frac{2}{\pi} \sqrt{1-x^2} \mathbf{1}_{|x| \leq 1}$ quand $n \rightarrow \infty$, c’est-à-dire que pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée on a :

$$\mathbb{E} \left[\int_{\mathbb{R}} g(x) \mu_n(\cdot, dx) \right] \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}} g(x) \rho(dx)$$

ce que l’on peut écrire en terme de fonctions de répartition :

$$\forall x \in \mathbb{R}, \mathbb{E}[\sigma_n(\cdot, x)] \xrightarrow{n \rightarrow +\infty} \int_{-\infty}^x \rho(u) du.$$

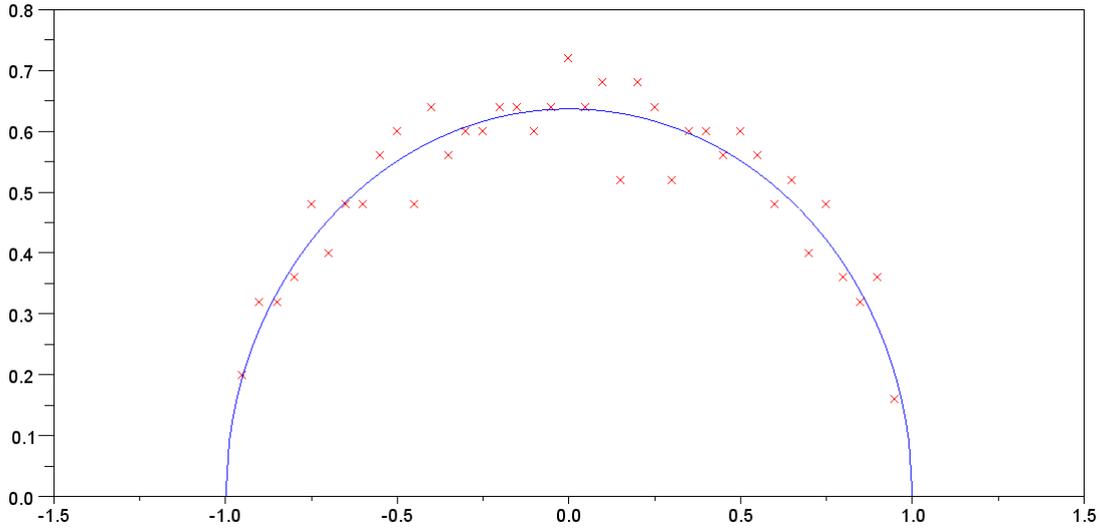


FIGURE 3 – Répartition des valeurs propres pour une matrice GOE de taille 500 et un pas de 0.05

On a alors :

$$\forall h \text{ petit}, \forall x, \frac{1}{h} \mathbb{E} \left[\frac{1}{n} \#\{\lambda \in \text{Sp } M/x < \lambda \leq x + h\} \right] \xrightarrow{n \rightarrow +\infty} \rho(x)$$

En pratique, le phénomène est *ergodique* : plus la taille de la matrice augmente et plus le spectre d'une matrice représente une grande moyenne d'ensemble de spectres. Cela nous permet d'enlever le signe espérance pour obtenir :

$$\forall h \text{ petit}, \forall x, \frac{1}{h} \frac{1}{n} \#\{\lambda \in \text{Sp } M/x < \lambda \leq x + h\} \xrightarrow{n \rightarrow +\infty} \rho(x).$$

Ainsi, pour vérifier la loi du demi-cercle, il suffit d'extraire le spectre d'une seule matrice de plus en plus grande et de vérifier que l'histogramme, tracé à pas constant, se rapproche de plus en plus du demi-cercle : les figures 3,4 et 5 en sont des illustrations. Des explications sur la construction des histogrammes sont fournies dans l'annexe B.1.

Note. Le demi-cercle est une densité asymptotique : la véritable forme du spectre est en fait quelque chose qui ressemble à un demi-cercle sur $[-1, 1]$: entre autres les branches du « demi-cercle » ne sont pas verticales et les valeurs propres ne sont pas bornées⁶. La figure 6 est une illustration de ce phénomène.

2 Distribution des valeurs propres : cas uniformément corrélé

Tout d'abord, dans toute la suite on supposera que les coefficients diagonaux sont nuls : leurs contributions aux valeurs propres sont négligeables devant celles des coef-

6. Mehta a calculé « explicitement » les densités non-asymptotiques dans [1].

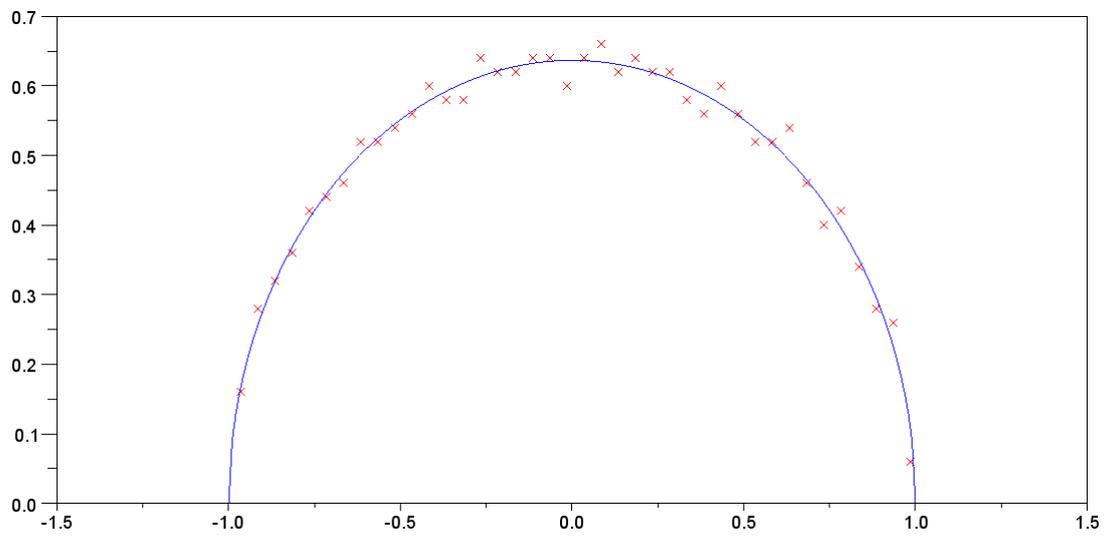


FIGURE 4 – Répartition des valeurs propres pour une matrice GOE de taille 1000 et un pas de 0.05

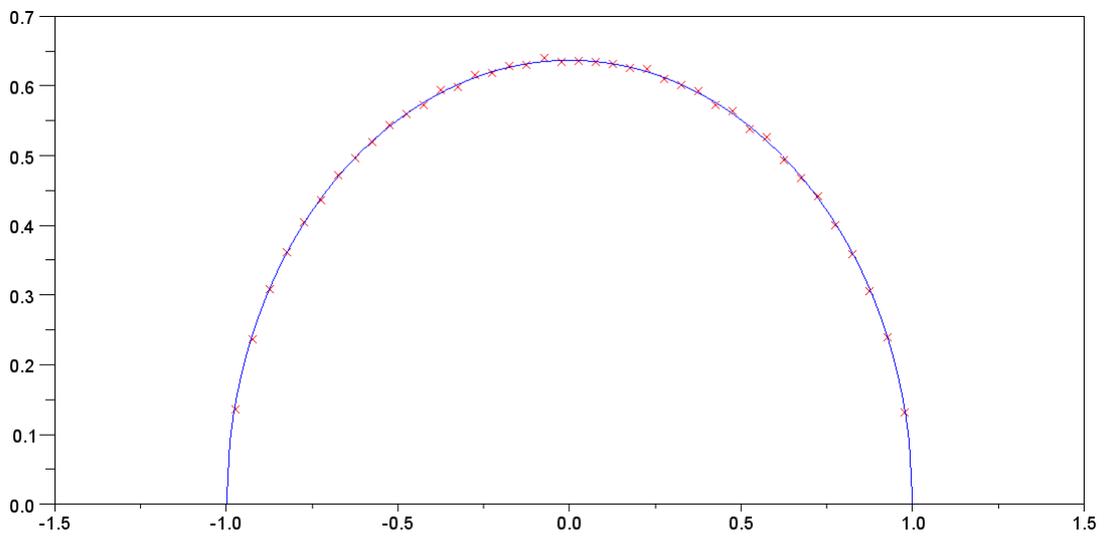


FIGURE 5 – Répartition des valeurs propres pour une matrice GOE de taille 10^4 et un pas de 0.05

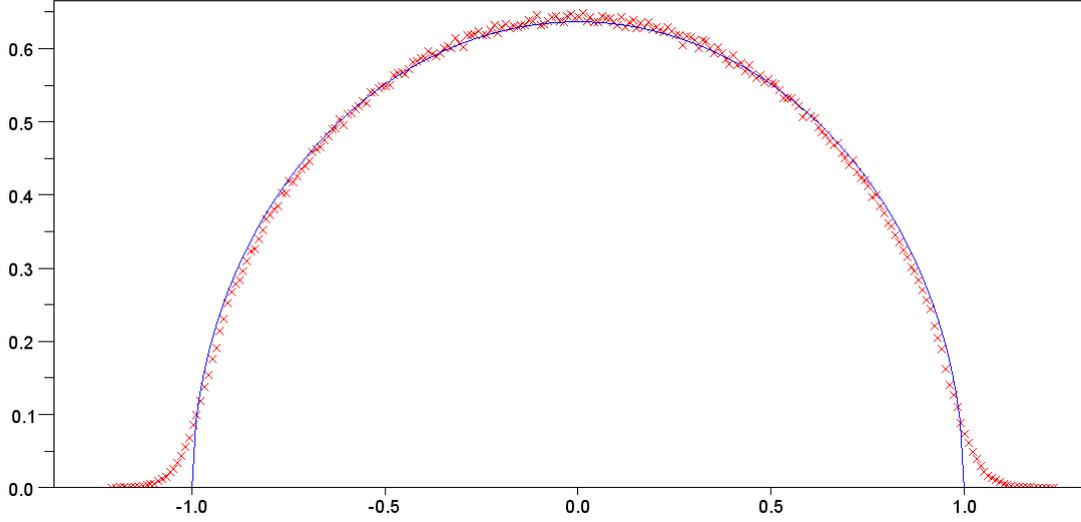


FIGURE 6 – Répartition des valeurs propres pour 10^5 matrices GOE de taille 30

ficients strictement sur-diagonaux ($\mathcal{O}(n)$ contre $\mathcal{O}(n^2)$), coefficients que l'on appellera donc désormais simplement « coefficients sur-diagonaux ». En outre, on enlèvera les coefficients diagonaux du vech qui sera alors de taille $N_p = \frac{n(n-1)}{2}$.

Un premier cas que l'on sait traiter quand on sort du GOE est le cas où tous les coefficients sont corrélés de la façon suivante :

$$M_i = \sqrt{1 - \rho}\sigma X_i + \sqrt{\rho}\sigma Y$$

où $X_i \sim Y \sim \mathcal{N}(0, 1)$ et où les X_i sont indépendants entre eux et avec Y . On remarque qu'il est aisé d'effectuer des tirages de matrices de ce type à partir de notre programme qui simule une matrice GOE.

La première chose à faire est de vérifier que l'on obtient bien la corrélation attendue.

$$\begin{aligned} \mathbb{E}[M_i M_j] &= \sigma^2 \left[(1 - \rho)\mathbb{E}[X_i X_j] + \rho\mathbb{E}[Y^2] + \sqrt{\rho(1 - \rho)}(\mathbb{E}[X_i Y] + \mathbb{E}[Y X_j]) \right] \\ &\stackrel{\perp}{=} \sigma^2 \rho \mathbb{E}[Y^2] \\ \mathbb{E}[M_i M_j] &= \sigma^2 \rho \end{aligned}$$

et on a donc bien $\rho(M_i, M_j) = \frac{\mathbb{E}[M_i M_j] - 0^2}{\sqrt{\sigma^2}\sqrt{\sigma^2}} = \rho$.

Ensuite, on peut regarder ce que l'on va obtenir. Sur la figure 7 s'aperçoit qu'il y a un pic (« glitch ») isolé, et sur la figure 8 on peut voir que l'on aura *a priori* un demi-cercle de rayon $\sqrt{1 - \rho}$. Ce « glitch » n'est pas une erreur de calcul ou d'approximation de la part de Scilab : en effet, si l'on suppose que toutes les valeurs propres ont la même loi, d'après [2] on peut supposer qu'elles se répartissent sur un demi-cercle de rayon a (note⁷). Ainsi, $\mathbb{E}[\lambda^2] = \frac{a^2}{4}$ et $\mathbb{E}[\text{tr}(M^2)] = \frac{n}{4}$: on trouve $a = 1$ ce qui contredit les résultats de la figure 8. On doit donc bien prendre en compte le « glitch ».

7. Arnold dit : « in general, the eigenvalues of a large [real symmetric] matrix will be approximately

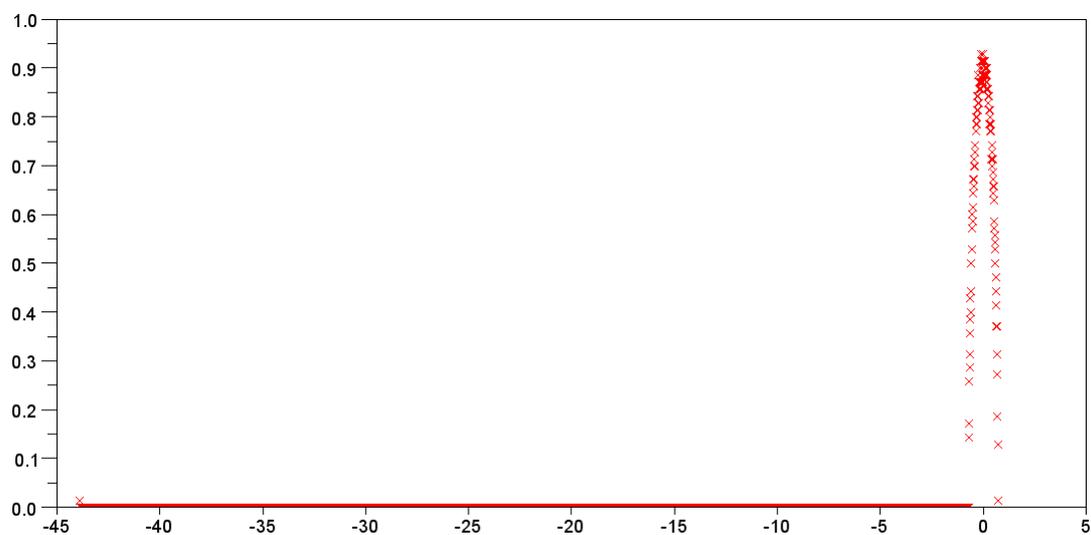


FIGURE 7 – Exemple de glitch (regarder tout à gauche) pour une matrice de taille 7000 avec $\rho = 0.5$

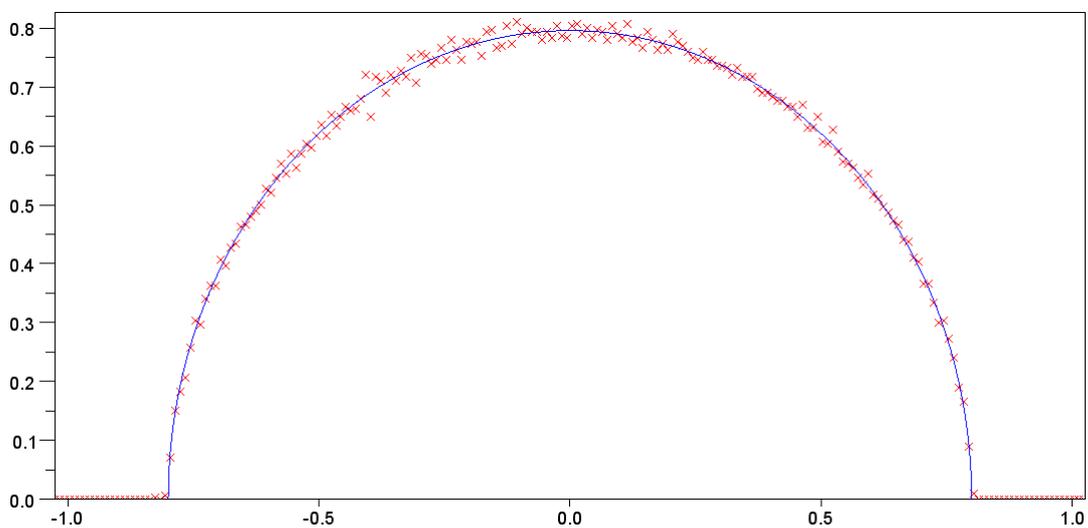


FIGURE 8 – Répartition des valeurs propres pour 10 matrices de taille 3000 avec $\rho = 0.36$

Il se trouve que l'on sait exactement où se trouve cette grande valeur propre : d'après [3], on peut dire qu'elle apparait quand $|M_0| > J$ et elle vaut alors :

$$\lambda_g = M_0 + \frac{J^2}{M_0}$$

où :

$$\begin{aligned} \frac{M_0}{n} &= \sqrt{\rho}\sigma Y \\ \frac{J^2}{n} &= (1 - \rho)\sigma^2 \end{aligned}$$

La condition $|M_0| > J$ s'écrit donc $|Y| > \sqrt{\frac{1-\rho}{n\rho}}$ et la valeur propre devient :

$$\lambda_g = n\sqrt{\rho}\sigma Y + \frac{(1 - \rho)\sigma^2}{\sqrt{\rho}\sigma Y} = \frac{\sqrt{n\rho}}{2} Y + \frac{1 - \rho}{2\sqrt{n\rho}Y}$$

Étant donné que l'on s'intéresse à une réalisation donnée de la matrice, on peut dans un premier temps négliger la condition ainsi que le deuxième terme de λ_g ($n \rightarrow +\infty$). On a donc $\lambda_g = \frac{\sqrt{n\rho}}{2} Y$, et $\mathbb{E}[\text{tr}(M^2)] = \frac{n}{4} = (n - 1)\frac{a^2}{4} + \frac{n\rho}{4} = \frac{n}{4}(a^2 + \rho)$ d'où $a^2 = 1 - \rho$ CQFD.

On peut en fait se passer de l'espérance :

$$\text{tr}(M^2) - \lambda_g^2 = \sum_{\lambda \neq \lambda_g} \lambda^2 \stackrel{\text{(note 8)}}{=} n\mathbb{E}[\lambda^2] = \frac{na^2}{4}$$

Ainsi, le carré du rayon est distribué selon la loi de $\frac{4}{n}(\text{tr}(M^2) - \lambda_g^2)$. Il faudrait donc retrancher à $\text{tr}(M^2)$ le terme $\lambda_g = n\sqrt{\rho}\sigma Y$ au carré pour avoir une idée du rayon. Or, quand on est face à une réalisation d'une matrice on n'a pas accès au Y . Pour remédier à ce problème, on va dire que $\sqrt{\rho}\sigma Y$ est la *moyenne des coefficients* $M_i = \sqrt{1 - \rho}\sigma X_i + \sqrt{\rho}\sigma Y$ sur la réalisation de la matrice : on peut alors estimer cette moyenne par

$$\overline{M} = \frac{1}{N_p} \sum_{i=1}^{N_p} M_i$$

et l'on remplace donc λ_g^2 par $n^2\overline{M}^2$. Ainsi, on remplace $\text{tr}(M^2) - \lambda_g^2$ par $\text{tr}(M^2) - n^2\overline{M}^2$; en divisant par n^2 , le terme devient $\frac{1}{n^2}\text{tr}(M^2) - \overline{M}^2 = \frac{1}{n^2} \sum_{i \neq j} M_{i,j}^2 - \left(\frac{1}{N_p} \sum_{i=1}^{N_p} M_i\right)^2$. On réécrit la première somme :

$$\frac{1}{n^2} \sum_{i \neq j} M_{i,j}^2 = \frac{1}{n^2} \sum_{i > j} M_{i,j}^2 + \frac{1}{n^2} \sum_{i < j} M_{i,j}^2 = \frac{2}{n^2} \sum_{i < j} M_{i,j}^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} M_i^2$$

distributed according to the semicircle law, if, roughly speaking, all the $u_{i,j}$'s ($i \geq j$) are of the same order of magnitude and are "independent" of each other », ce qui est le cas ici.

8. Les valeurs propres ne sont ici pas indépendantes, mais l'ergodicité nous permet de conserver la loi des grands nombres.

On a finalement $\frac{1}{n^2} \text{tr}(M^2) - \bar{M}^2 = \frac{1}{N_p} \sum M_i^2 - \left(\frac{1}{N_p} \sum M_i \right)^2$: c'est exactement $\bar{\sigma}^2$, l'estimateur de la variance des coefficients sur-diagonaux de la matrice⁹ ! Le carré du rayon du demi-cercle est donc distribué selon la loi de $\frac{4}{n} n^2 \bar{\sigma}^2 = 4n \bar{\sigma}^2 = \frac{\bar{\sigma}^2}{\sigma^2}$; on obtient ainsi le résultat suivant.

Conjecture (R. Delannay, G. Le Caër). *Soit M est une matrice aléatoire symétrique réelle dont la taille tend vers l'infini telle que :*

- $\mathbb{E}[M_i] = 0$
- $\mathbb{E}[M_i^2] = \sigma^2$

L'histogramme des valeurs propres de $M(\omega)$ est un demi-cercle de rayon $\frac{\bar{\sigma}(\omega)}{\sigma}$ et l'histogramme de l'ensemble des valeurs propres des $M(\omega)$ est donné par :

$$D(x) = \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u-x^2}}{u} \mathbb{P}_{\frac{\bar{\sigma}^2}{\sigma^2}}(du).$$

L'intégrale intervient uniquement pour faire la « somme » des demi-cercles normalisés (par $\frac{2}{\pi u}$, u désignant le carré du rayon) obtenus. Revenons maintenant à notre problème de corrélation uniforme : on veut trouver la loi de $\frac{\bar{\sigma}^2}{\sigma^2}$. On a :

$$\begin{aligned} \bar{\sigma}^2 &= \frac{1}{N_p} \sum (M_i - \bar{M})^2 \\ &= \frac{\sigma^2}{N_p} \sum (\sqrt{1-\rho} X_i + \sqrt{\rho} Y - \sqrt{1-\rho} \bar{X} - \sqrt{\rho} \bar{Y})^2 \\ \bar{\sigma}^2 &= \frac{\sigma^2(1-\rho)}{N_p} \sum (X_i - \bar{X})^2 \end{aligned}$$

En posant $Z = \sum (X_i - \bar{X})^2$, comme les X_i sont indépendantes il est connu que Z suit une loi $\chi_{N_p-1}^2$: une démonstration de ce résultat est donnée dans l'annexe D. On a donc $\mathbb{E}[Z] = N_p - 1$, $\text{Var} Z = 2(N_p - 1)$ et :

$$\begin{aligned} \mathbb{E}\left[\frac{\bar{\sigma}^2}{\sigma^2}\right] &= \frac{1-\rho}{N_p} (N_p - 1) = (1-\rho) \\ \text{Var}\left(\frac{\bar{\sigma}^2}{\sigma^2}\right) &= \frac{(1-\rho)^2}{N_p^2} 2(N_p - 1) = 2 \frac{(1-\rho)^2}{N_p} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

Ainsi, d'après l'inégalité de Tchebychev :

$$\forall \varepsilon > 0, \mathbb{P}\left(\left|\frac{\bar{\sigma}^2}{\sigma^2} - (1-\rho)\right| > \varepsilon\right) \leq \frac{\text{Var}\left(\frac{\bar{\sigma}^2}{\sigma^2}\right)}{\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0.$$

La suite $\left(\frac{\bar{\sigma}^2}{\sigma^2}\right)_n$ converge donc en probabilité vers sa moyenne $1-\rho$, et donc en loi. On a donc :

$$\int_{x^2}^{+\infty} \frac{\sqrt{u-x^2}}{u} \mathbb{P}_{\frac{\bar{\sigma}^2}{\sigma^2}}(du) \xrightarrow{n \rightarrow +\infty} \int_{x^2}^{+\infty} \frac{\sqrt{u-x^2}}{u} \delta_{1-\rho}(du) = \frac{\sqrt{1-\rho-x^2}}{1-\rho} \mathbf{1}_{x^2 \leq 1-\rho}$$

9. Cet estimateur est biaisé : il faudrait remplacer $\frac{1}{N_p}$ par $\frac{1}{N_p-1}$ pour que son espérance vaille exactement $\text{Var}(X_i)$. Comme $N_p \rightarrow +\infty$, on peut négliger le -1 .

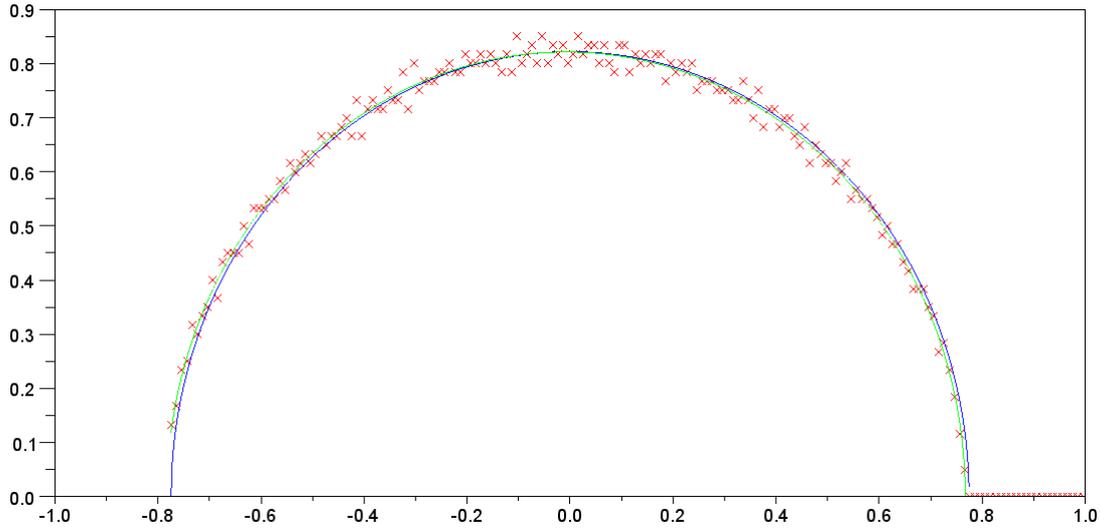


FIGURE 9 – Décalage des valeurs propres pour une matrice 6000 avec $\rho = 0.4$

D'où :

$$D(x) = \frac{2}{\pi(1-\rho)} \sqrt{1-\rho-x^2} \mathbf{1}_{x^2 \leq 1-\rho}$$

On retrouve bien le demi-cercle de rayon $\sqrt{1-\rho}$!

Note. Ce résultat ne dépend pas de la loi de Y .

La figure 9 montre que pour un évènement donné, les valeurs propres peuvent être légèrement décalées du demi-cercle théorique. Sur la figure, sur laquelle on a ajusté un demi-cercle par moindres carrés, le décalage est de $\delta = -0.0079$: ce décalage provient en fait de la valeur propre glitch. En effet, $\text{tr } M = 0 = \sum \lambda + \lambda_g$ donc en divisant par n , $0 = \delta + \frac{1}{n}(M_0 + \frac{J^2}{M_0})$. Remarquons que la condition $|M_0| > J$ est nécessairement satisfaite s'il y a un décalage, faute de quoi on aurait $\delta = 0$ (toutes les valeurs propres seraient sur le demi-cercle). Ainsi, on a $\delta = \frac{-1}{n} \left(M_0 + \frac{J^2}{M_0} \right)$. Dans la matrice de la figure 9, le Y valait 1.9138 : le décalage théorique est donc de -0.0078 . On retrouve bien la valeur ajustée!

En particulier, il faut remarquer que la valeur propre glitch joue un *énorme* rôle dans le moment d'ordre 2 : si $\lambda \neq \lambda_g$, alors $\mathbb{E}[\lambda^2] = \frac{1-\rho}{4}$ car elle se répartit sur le demi-cercle de rayon $\sqrt{1-\rho}$. Ainsi, $\mathbb{E}[\text{tr } M^2] = \frac{n}{4} = \sum \mathbb{E}[\lambda^2] + \mathbb{E}[\lambda_g^2] = n \frac{1-\rho}{4} + \mathbb{E}[\lambda_g^2]$: on retrouve le fait que le moment d'ordre 2 de λ_g vaut $\frac{n\rho}{4}$ (note¹⁰), ce qui est *considérable* — et tend même vers l'infini! — pour une valeur propre de poids $\frac{1}{n}$ dans l'histogramme!

10. On avait déjà ce résultat avec la formule $\lambda_g = \frac{\sqrt{n\rho}}{2} Y$.

3 Distribution des valeurs propres : corrélation par blocs

On entre ici dans le cœur du stage. Pour la corrélation uniforme, l'ordre des coefficients dans le vech n'importait pas : ici en revanche, la matrice de corrélation va être *diagonale par blocs* et l'ordre dans le vech est donc important. Lorsque l'on fait des tirages, on le choisit en fait aléatoirement au début et il reste le même pour toutes les matrices du tirage.

On suppose que tous les blocs sont *identiques* ; par exemple, une matrice de corrélation que l'on considère est :

$$\begin{pmatrix} 1 & \rho & \rho & & & \\ \rho & 1 & \rho & & & \\ \rho & \rho & 1 & & & \\ & & & 1 & \rho & \rho \\ & & & \rho & 1 & \rho \\ & & & \rho & \rho & 1 \end{pmatrix}$$

3.1 Nombre de blocs constant

On note $m \geq 2$ le nombre de blocs : on suppose que m est *constant* (i.e. $\frac{\partial m}{\partial n} = 0$). De plus, on note $b(i)$ l'indice du bloc associé à l'indice i du vech :

$$M_i = \sqrt{1 - \rho}\sigma X_i + \sqrt{\rho}\sigma Y_{b(i)}$$

La conjecture dit que l'on « somme » les demi-cercles. Dans notre cas, on peut explicitement observer le phénomène : c'est ce qu'illustre la figure 10. On voit bien qu'il suffit de connaître la loi du rayon du demi-cercle, donnée par $\frac{\bar{\sigma}}{\sigma}$, pour déterminer la forme de l'histogramme global. On remarquera que le phénomène n'est cette fois plus ergodique : le spectre d'une grosse matrice ne rend pas compte d'une moyenne d'ensemble de spectres.

En suivant la conjecture, on désirerait donc trouver la loi de $\frac{\bar{\sigma}^2}{\sigma^2}$. On a :

$$\bar{\sigma}^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\sqrt{1 - \rho}\sigma X_i + \sqrt{\rho}\sigma Y_{b(i)} \right)^2 - \left[\frac{1}{N_p} \sum_{i=1}^{N_p} \left(\sqrt{1 - \rho}\sigma X_i + \sqrt{\rho}\sigma Y_{b(i)} \right) \right]^2$$

puis :

$$\begin{aligned} \frac{\bar{\sigma}^2}{\sigma^2} &= \frac{1}{N_p} \sum_{i=1}^{N_p} \left[(1 - \rho)X_i^2 + \rho Y_{b(i)}^2 + 2\sqrt{\rho(1 - \rho)}X_i Y_{b(i)} \right] \\ &\quad - \frac{1}{N_p^2} \left(\sqrt{1 - \rho} \sum_{i=1}^{N_p} X_i + \sqrt{\rho} \sum_{i=1}^{N_p} Y_{b(i)} \right)^2 \end{aligned}$$

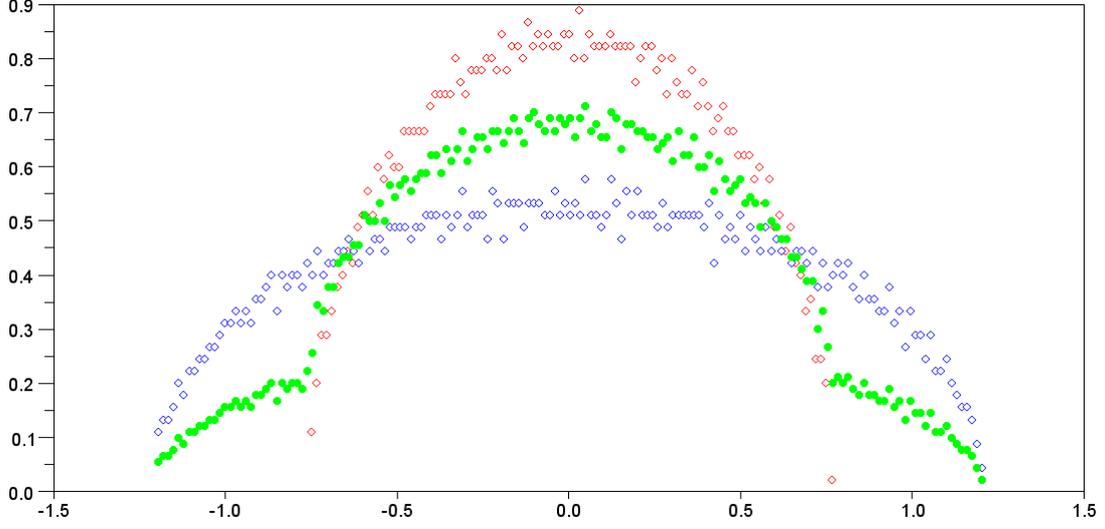


FIGURE 10 – Somme de deux spectres

On introduit les notations suivantes :

$$\bar{X}^j = \frac{1}{N_p} \sum_{i=1}^{N_p} X_i^j$$

$$\bar{Y}^j = \frac{1}{m} \sum_{i_b=1}^m Y_{i_b}^j$$

On continue :

$$\frac{\bar{\sigma}^2}{\sigma^2} = (1 - \rho) \bar{X}^2 + \frac{\rho}{N_p} \sum_{i=1}^{N_p} Y_{b(i)}^2 + 2 \frac{\sqrt{\rho(1-\rho)}}{N_p} \sum_{i=1}^{N_p} X_i Y_{b(i)}$$

$$- (1 - \rho) \bar{X}^2 - \frac{\rho}{N_p^2} \left(\sum_{i=1}^{N_p} Y_{b(i)} \right)^2 - 2 \frac{\sqrt{\rho(1-\rho)}}{N_p^2} \left(\sum_{i=1}^{N_p} X_i \right) \left(\sum_{i=1}^{N_p} Y_{b(i)} \right)$$

Les m blocs sont de tailles égales, donc les blocs sont de taille $\frac{N_p}{m}$. On a alors :

$$\frac{\bar{\sigma}^2}{\sigma^2} = (1 - \rho) \bar{X}^2 + \frac{\rho}{N_p} \sum_{i_b=1}^m \frac{N_p}{m} Y_{i_b}^2 + 2 \frac{\sqrt{\rho(1-\rho)}}{N_p} \sum_{i_b=1}^m Y_{i_b} \sum_{\substack{i \\ b(i)=i_b}} X_i$$

$$- (1 - \rho) \bar{X}^2 - \frac{\rho}{N_p^2} \left(\sum_{i_b=1}^m \frac{N_p}{m} Y_{i_b} \right)^2 - 2 \frac{\sqrt{\rho(1-\rho)}}{N_p} \bar{X} \left(\sum_{i_b=1}^m \frac{N_p}{m} Y_{i_b} \right)$$

En simplifiant les N_p et en réorganisant les termes on obtient :

$$\begin{aligned} \frac{\bar{\sigma}^2}{\sigma^2} &= (1 - \rho)\overline{X^2} - (1 - \rho)\bar{X}^2 + \rho\overline{Y^2} - \rho\bar{Y}^2 \\ &\quad + 2\sqrt{\rho(1 - \rho)} \sum_{i_b=1}^m Y_{i_b} \left(\frac{1}{N_p} \sum_{b(i)=i_b} X_i \right) - 2\sqrt{\rho(1 - \rho)}\bar{X}\bar{Y} \quad (9) \end{aligned}$$

D'après la loi des grands nombres, on a quand $n \rightarrow +\infty$:

$$\begin{aligned} \bar{X} &\rightarrow 0 \\ \overline{X^2} &\rightarrow 1 \end{aligned}$$

De plus, comme $N_p \geq \#\{i/b(i) = i_b\} (= \frac{N_p}{m})$ on a également :

$$\frac{1}{N_p} \sum_{b(i)=i_b} X_i \xrightarrow{n \rightarrow +\infty} 0$$

Enfin, comme m reste fixe quand n varie et comme Y_{i_b} est fini on peut en déduire que :

$$\sum_{i_b=1}^m Y_{i_b} \left(\frac{1}{N_p} \sum_{b(i)=i_b} X_i \right) \xrightarrow{n \rightarrow +\infty} 0$$

L'équation (9) devient finalement donc quand n est grand :

$$\frac{\bar{\sigma}^2}{\sigma^2} = 1 - \rho + \rho(\overline{Y^2} - \bar{Y}^2) \quad (10)$$

La variable Y est une $\mathcal{N}(0, 1)$, donc comme on l'a déjà vu la variable $Z = m(\overline{Y^2} - \bar{Y}^2)$ suit une loi χ_{m-1}^2 . On a donc :

$$\frac{\bar{\sigma}^2}{\sigma^2} \sim (1 - \rho) + \frac{\rho}{m} \chi_{m-1}^2$$

$\frac{\bar{\sigma}^2}{\sigma^2}$ est donc à densité et cette densité vaut :

$$u \mapsto \frac{1}{\Gamma(\frac{m-1}{2})} \left(\frac{m}{2\rho} \right)^{\frac{m-1}{2}} [u - (1 - \rho)]^{\frac{m-3}{2}} e^{-m \frac{u-(1-\rho)}{2\rho}} \mathbf{1}_{u \geq 1-\rho}$$

La densité des valeurs propres sur le demi-cercle est donc :

$$\begin{aligned} D(x) &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \mathbb{P}_{\frac{\sigma^2}{2}}(du) \\ &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \frac{1}{\Gamma(\frac{m-1}{2})} \left(\frac{m}{2\rho} \right)^{\frac{m-1}{2}} [u - (1 - \rho)]^{\frac{m-3}{2}} e^{-m \frac{u-(1-\rho)}{2\rho}} \mathbf{1}_{u \geq 1-\rho} du \\ D(x) &= \frac{2}{\pi \Gamma(\frac{m-1}{2})} \left(\frac{m}{2\rho} \right)^{\frac{m-1}{2}} \int_{\max(1-\rho, x^2)}^{+\infty} \frac{\sqrt{u - x^2}}{u} [u - (1 - \rho)]^{\frac{m-3}{2}} e^{-m \frac{u-(1-\rho)}{2\rho}} du \end{aligned}$$

Les figures 11, 12 et 13 confirment cette forme; les détails des programmes sont fournis dans l'annexe A.2.

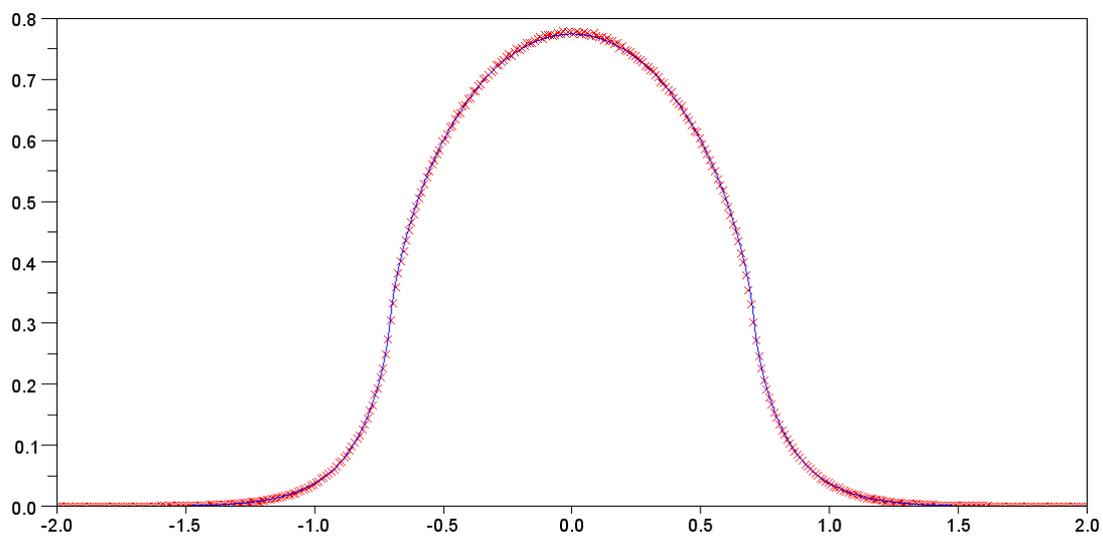


FIGURE 11 – Répartition des valeurs propres pour 10^4 matrices de taille 500 avec une corrélation à 2 blocs identiques et $\rho = 0.5$

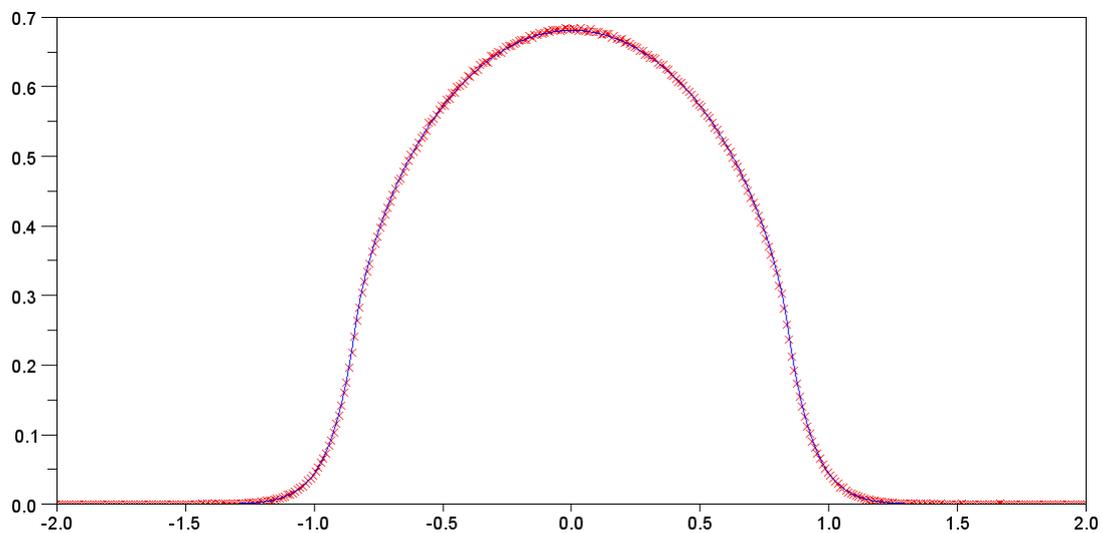


FIGURE 12 – Répartition des valeurs propres pour 10^4 matrices de taille 501 avec une corrélation à 3 blocs identiques et $\rho = 0.3$

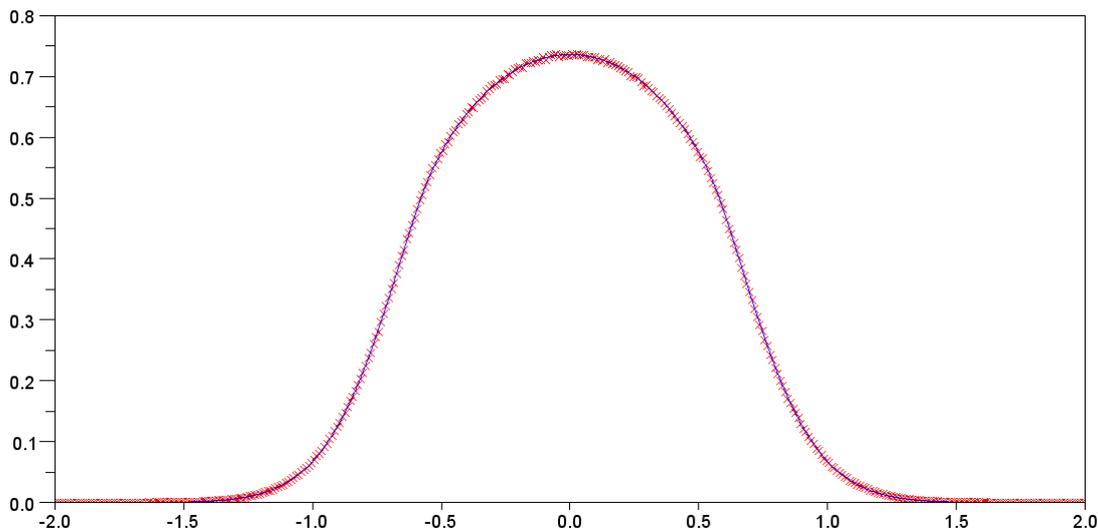


FIGURE 13 – Répartition des valeurs propres pour 10^4 matrices de taille 500 avec une corrélation à 5 blocs identiques et $\rho = 0.7$

Note. Ce résultat dépend de la loi des termes d'ajout Y_{i_b} : en effet, on a utilisé le fait qu'ils sont gaussiens pour dire que la variable $m(\overline{Y^2} - \overline{Y}^2)$ suit une loi du χ^2 . On peut donc se demander quels histogrammes on obtient si l'on change la loi des Y_{i_b} : cela sortant d'un contexte physique, j'ai mis quelques résultats (où les Y_{i_b} suivent des lois uniformes) dans l'annexe E.

3.2 Taille des blocs constante

On suppose maintenant que m varie avec n , et plus précisément que $\frac{N_p}{m}$ (la taille des blocs) est constant avec n . On peut reprendre tout le raisonnement précédent jusqu'à l'équation (9) :

$$\begin{aligned} \frac{\overline{\sigma^2}}{\sigma^2} &= (1 - \rho)\overline{X^2} - (1 - \rho)\overline{X}^2 + \rho\overline{Y^2} - \rho\overline{Y}^2 \\ &\quad + 2\sqrt{\rho(1 - \rho)} \sum_{i_b=1}^m Y_{i_b} \left(\frac{1}{N_p} \sum_{b(i)=i_b} X_i \right) - 2\sqrt{\rho(1 - \rho)}\overline{X}\overline{Y} \end{aligned}$$

On a toujours $\overline{X} \xrightarrow[\text{LGN}]{n \rightarrow +\infty} 0$, $\overline{X^2} \xrightarrow[\text{LGN}]{n \rightarrow +\infty} 1$ et on a cette fois-ci également les mêmes convergences pour Y car $m \xrightarrow{n \rightarrow +\infty} +\infty$. En revanche, on ne peut plus cuisiner le terme $\sum_{i_b=1}^m Y_{i_b} \left(\frac{1}{N_p} \sum_{b(i)=i_b} X_i \right)$ comme avant justement car la somme sur i_b devient de plus en plus grande. Pour remédier à ce problème, on pose :

$$Z_{i_b} = Y_{i_b} \left(\sum_{b(i)=i_b} X_i \right)$$

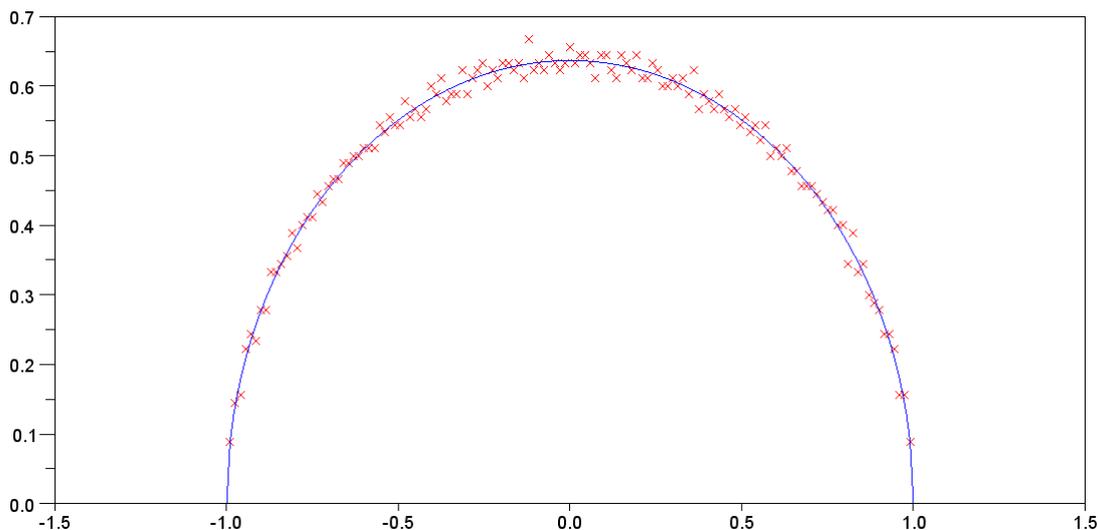


FIGURE 14 – Répartition des valeurs propres pour une matrice de taille 6000, avec une corrélation par blocs identiques de taille 2 et $\rho = 0.5$

Comme les variables X_i sont i.i.d. ainsi que les Y_{i_b} , comme les X_i et les Y_{i_b} sont indépendantes entre elles alors les variables Z_{i_b} sont elles aussi i.i.d. et d'espérance nulle. Ainsi,

$$\bar{Z} \xrightarrow[\text{LGN}]{n \rightarrow +\infty} 0$$

et comme $N_p \geq m$ on en déduit que $\frac{1}{N_p} \sum_{i_b=1}^m Z_{i_b}$ tend aussi vers 0. L'équation (9) devient donc à la limite :

$$\frac{\bar{\sigma}^2}{\sigma^2} = (1 - \rho) - 0 + \rho - 0 + 0 - 0$$

donc $\frac{\bar{\sigma}^2}{\sigma^2} = 1$! Ainsi, la densité des valeurs propres devient :

$$\begin{aligned} D(x) &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \mathbb{P}_{\frac{\sigma^2}{\sigma^2}}(du) \\ &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \delta_1(du) \\ D(x) &= \frac{2}{\pi} \sqrt{1 - x^2} \mathbf{1}_{x^2 \leq 1} \end{aligned}$$

On retrouve donc le demi-cercle de rayon 1 : on observe bien cela sur les figures 14, 15 et 16. On pouvait s'y attendre, car si la taille des blocs est constante alors elle devient négligeable quand n devient infini et « à la limite on ne distingue plus que les 1 de la diagonale » : on se retrouve alors dans le cas GOE, d'où le demi-cercle. En outre, le phénomène redevient ergodique.

Note. Le résultat ne dépend pas de la loi des termes d'ajout Y_{i_b} (cependant, ils doivent rester i.i.d.).

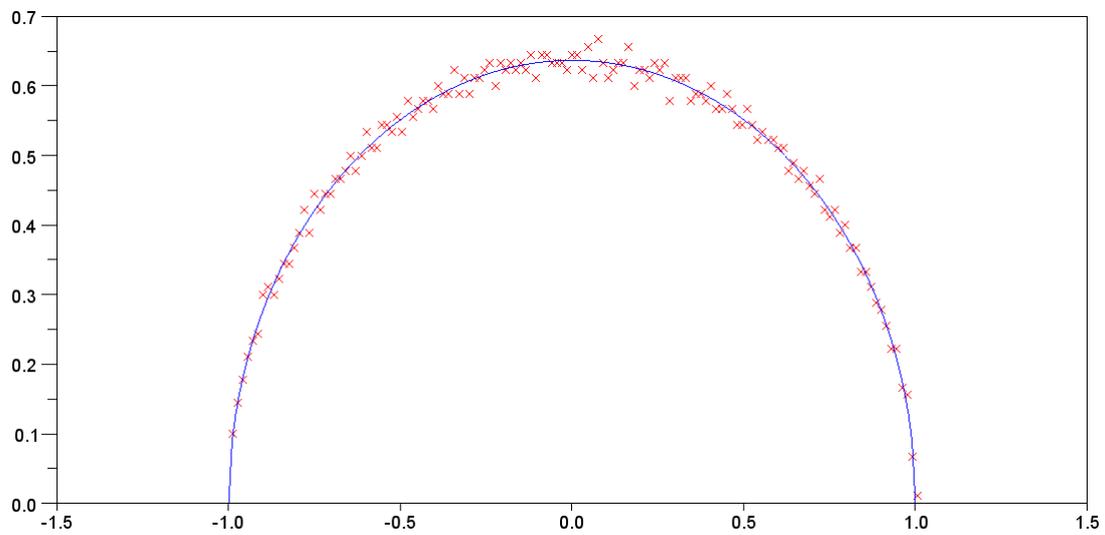


FIGURE 15 – Répartition des valeurs propres pour une matrice de taille 6000, avec une corrélation par blocs identiques de taille 3 et $\rho = 0.3$

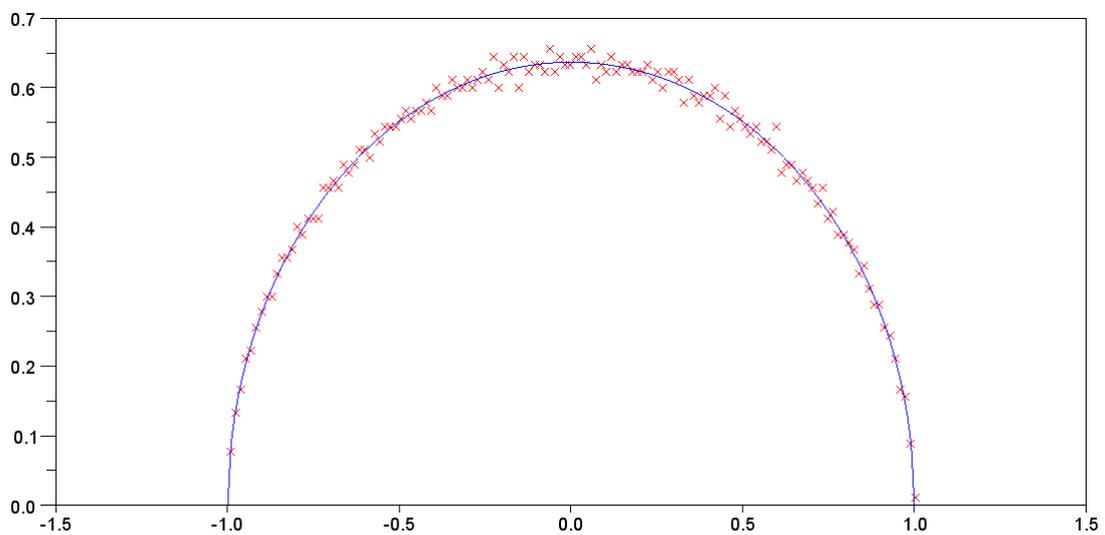


FIGURE 16 – Répartition des valeurs propres pour une matrice de taille 6000, avec une corrélation par blocs identiques de taille 5 et $\rho = 0.7$

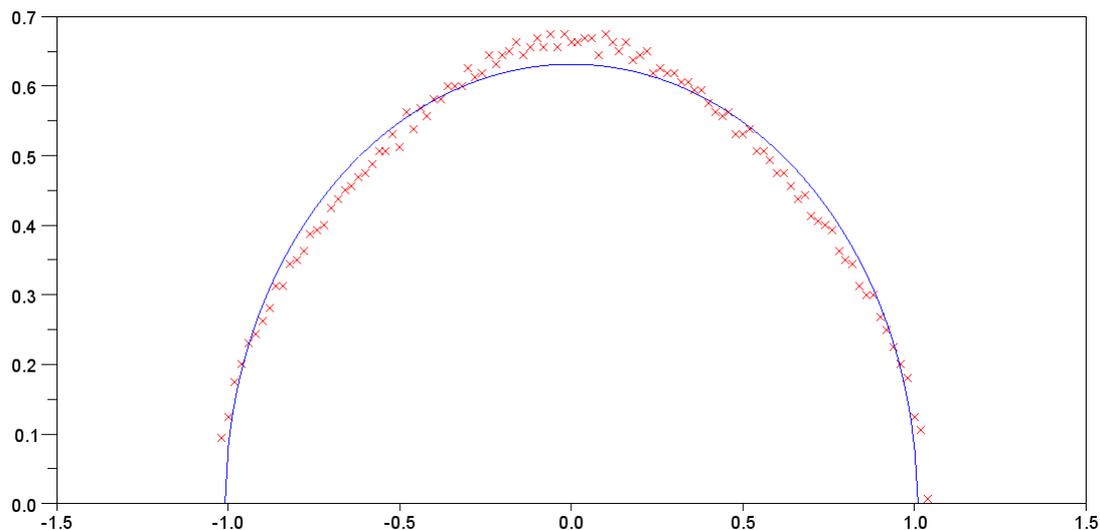


FIGURE 17 – Répartition des valeurs propres pour une matrice de taille 8000, avec une matrice de corrélation avec des blocs identiques de taille 2 et $\rho = 0.5$

3.3 Cas pathologiques

On pourrait croire que tous les cas de corrélations par blocs égaux sont traités. En outre, l'ordre du vecteur n'est *a priori* pas rentré en compte : et pour cause, on le choisit aléatoirement au début de chaque série de tirages. Ainsi, si l'on corrèle les coefficients sur-diagonaux deux par deux verticalement (en prenant le coefficient du haut de la colonne suivante si besoin), on obtient l'histogramme 17 ; en les corrélant quatre par quatre verticalement (même remarque), on obtient l'histogramme 18. On constate que le modèle n'est plus valide et en particulier le résultat de [2] puisque les valeurs propres d'une matrice déterministe ne se répartissent plus sur un demi-cercle.

Dans le cadre plus général de corrélations avec des blocs quelconques, on rencontre également des problèmes pour des corrélations par exemple de type Toeplitz (corrélations par diagonales) ou bien de type Hankel (corrélations par antidiagonales).

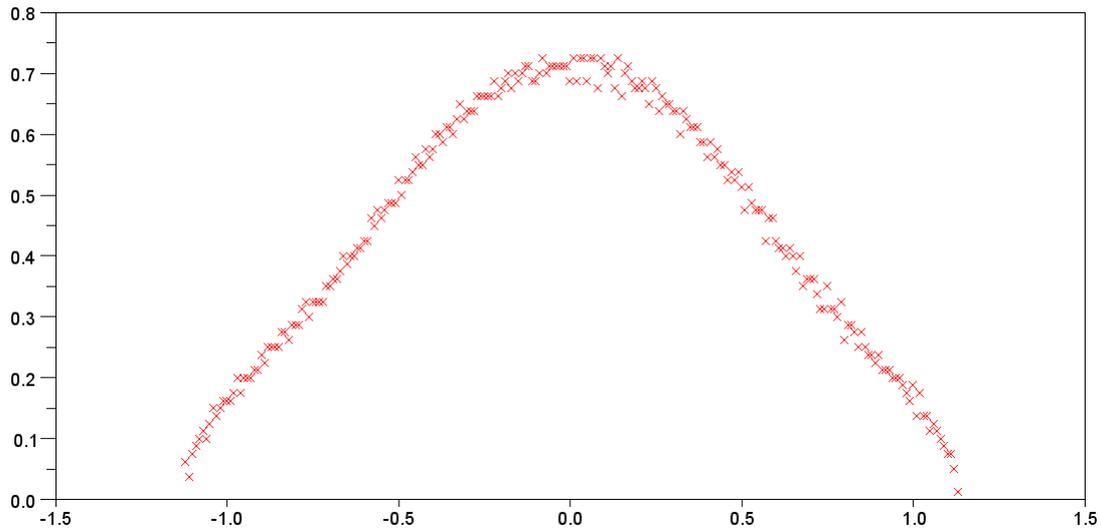


FIGURE 18 – Répartition des valeurs propres pour une matrice de taille 8000, avec une matrice de corrélation avec des blocs identiques de taille 4 et $\rho = 0.5$

Conclusion

On a traité les cas de corrélations par blocs pour des blocs *identiques*; cependant, les cas où les blocs sont de tailles différentes (par exemple $[1, 2, 3, \dots, n - 1]$ pour le cas Toeplitz) ne sont pas traités ici. Avec une approche différente, on peut résoudre des cas du type deux blocs de tailles différentes, et même des cas plus généraux avec un nombre de blocs quelconque et des rapports de taille constant (par exemple $[\frac{1}{2}, \frac{1}{3}, \frac{1}{6}]$) : cela revient alors à trouver la loi de la somme de variables gamma, bien sûr quand les variables corrélatives (les Y_{i_b}) sont gaussiennes.

Il est important de préciser que l'origine des « cas pathologiques » (sous-section 3.3) est pour l'instant mystérieuse; en particulier, on n'a pas de critère simple pour dire si le modèle va marcher ou non. Peut-être le critère d'Arnold de [2] suffit, mais il n'est à première vue pas très simple à manier.

Références

- [1] M.L. Mehta, *Random matrices* (third edition). Elsevier Academic Press, 2004.
- [2] L. Arnold, *Deterministic Version of Wigners's Semicircle Law for the Distribution of Matrix Eigenvalues*. Linear algebra and its applications 13,185–199, 1976.
- [3] S.F. Edwards et R.C. Jones, *The eigenvalue spectrum of a large symmetric random matrix*. Journal of Physics A, Math. Gen vol. 9 N° 10, 1976.
- [4] C. Hammond et S.J. Miller, *Distribution of Eigenvalues for the Ensemble of Real Symmetric Toeplitz Matrices*. Journal of Theoretical Probability, Vol. 18, N° 3, Juillet 2005.
- [5] W. Bryc, A. Dembo et T. Jiang, *Spectral mesure of large random Hankel, Markov and Toeplitz matrices*. The Annals of Probability, Vol. 34, N° 1, 1–38, 2006.
- [6] O. Friesen et M. Löwe, *The semicircle law for matrices with independent diagonals*. À paraître dans le Journal of Theoretical Probability, 2011.
- [7] O. Friesen et M. Löwe, *A phase transition for the limiting spectral density of random matrices*. Mai 2012.

A Création des différentes matrices

A.1 Cas GOE

```
function H=genmat(N)
    sD=1/sqrt(2*(N+1))
    sND=sD/sqrt(2)

    H=diag(grand(1,N,'nor',0,sD))

    for i=1:N-1 do
        ajout=grand(1,N-i,'nor',0,sND)
        H(i,i+1:N)=ajout
        H(i+1:N,i)=ajout'
    end
endfunction
```

A.2 Cas de corrélations par blocs identiques

Il serait facile de créer la matrice de corrélation puis de générer les variables en utilisant la décomposition de Choleski de la matrice de corrélation. Mais voilà, on arrive à une complexité en espace en $\mathcal{O}(n^4)$ (la matrice de corrélation étant de taille $\mathcal{O}(n^2)$). On va donc simplement générer la matrice GOE puis ajouter à certains coefficients de terme $Y_{b(i)}$ qui va corrélérer. Les coefficients qui sont corrélés entre eux sont choisis de manière aléatoire au début de chaque série de tirage : concrètement, chaque figure correspond à un ordre du vech différent. J'ai utilisé le programme suivant, qui peut générer une corrélations par blocs quelconque (les coefficients de corrélations ainsi que les tailles peuvent être différents), où :

- `cor` est le tableau des différents coefficients de corrélation ;
- `taille_blocs` est le nombre de coefficients qui sont corrélés avec le même coefficient de corrélation (la somme de ses coefficients vaut alors N_p) : la matrice de corrélation est donc $\text{diag}(J_1, \dots, J_m)$ où m est la longueur de `taille_blocs` (*i.e.* le nombre de blocs) et où chaque J_i est un bloc de corrélation uniforme avec un coefficient de corrélation égal à `cor.(i)` ;
- `indices_mat` fourni l'ordre des coefficients dans le vech : c'est une permutation des indices sur-diagonaux (les coefficients d'une matrice étant indexés de 1 à n^2).

```
function H=cor_blocs(n,cor,taille_blocs,indices_mat)
    sD=1/sqrt(2*n+2)
    sND=sD/sqrt(2)
    m=length(cor)

    H=zeros(n,n)
    gaussienne_correlative=sND*rand(1,nb,'normal')
    u=1;v=0
```

```

for i=1:m do
    lb=taille_blocs(i);rho=cor(i)
    v=u+lb-1
    r=sqrt(rho);rr=sqrt(1-rho)
    gaussienne=sND*rand(1,lb,'normal')
    ajout=rr*gaussienne+r*gaussienne_correlative(i)
    k=indices_mat(u:v)
    p=floor(k/n);q=k-p*n;kt=(q-1)*n+p+1
    // trouve les indices transposés correspondants
    H(k)=ajout
    H(kt)=ajout'
    u=v+1
end
endfunction

```

(On aurait aussi pu remplir la partie triangulaire supérieure puis additionner la transposée.) Le tableau `indices_mat` est généré grâce à la fonction suivante : on rappelle que les coefficients des matrices à une coordonnée sont indexés de haut en bas puis de gauche à droite.

```

function t=cree_indicesmat_aleat(n)
    Np=n*(n-1)/2
    t=zeros(Np,1)
    i=1
    for p=1:n-1 do
        t(i:i+p-1)=p*n+(1:p)'
        i=i+p
    end
    t=grand(1,'prm',t)
endfunction

```

Finalement, si l'on veut une matrice de corrélation avec des blocs identiques on a juste à appeler la fonction `cor_blocs` avec l'argument `rho*ones(1,m)` à la place de `cor` ainsi que `Np/m*ones(1,m)` à la place de `taille_blocs`. On remarque que l'on doit au préalable s'assurer que $\frac{Np}{m}$ est bien un entier !

Hélas, les problèmes ne s'arrêtent pas là : la densité théorique, quand elle est donnée sous la forme d'une intégrale, va être délicate à tracer sous Scilab. J'ai utilisé les instructions qui suivent.

```

function u=integrande(u,x)
    u=sqrt(u-x^2)*(u-(1-rho))^(nb-3)/2/(u*exp(nb/(2*rho)*(u-(1-rho))))
endfunction

```

```

function y=D(x)
    b=max(1-rho,x^2)
    y=2/%pi*(nb/(2*rho))^(nb-1)/2)/gamma((nb-1)/2)
    *intg(b,max(b+1,10),list(integrande,x))
endfunction

function y=D_tab(x)
    y=x
    for i=1:length(x) do
        y(i)=D(x(i))
    end
endfunction

```

B Graphiques

B.1 Histogramme

La fonction `histplot` de Scilab est très efficace pour créer des histogrammes, mais le problème est qu'elle ne renvoie aucune information à son propos. En fait, j'ai créé une fonction qui :

- renvoie l'histogramme des valeurs avec un pas donné en argument ; l'histogramme est normalisé, c'est-à-dire que les ordonnées sont données par :

$$\frac{1}{\#\text{échantillon} \times \text{pas}} \#\text{classe},$$

ce qui permet d'avoir une intégrale de 1 ;

- renvoie dans un fichier `.txt` les abscisses des milieux des classes ainsi que le pas et les ordonnées ;
- si la densité théorique d est fournie, le programme rajoute dans le fichier `.txt` les valeurs moyennes de la densité théorique sur les classes (ou les valeurs de la densité théorique aux milieux des classes si la primitive n'est pas calculable), données par :

$$\frac{1}{\text{pas}} \int_{x_i}^{x_{i+1}} d(x) dx.$$

Il faut prendre garde aux effets d'histogramme, c'est-à-dire qu'il ne faut pas trop tenir compte des points extrêmes si la courbe théorique est à support borné.

B.2 Ajustement par la méthode des moindres carrés

On utilise la fonction `leastsq` de Scilab. Par exemple, pour ajuster un seul demi-cercle on utilise le script suivant.

```

function y=demi_cercle(x,r,c)
//r est le rayon, c le centre

```

```

y=zeros(x)
ind=find(abs(x-c)<=r)
y(ind)=2/(%pi*r^2)*sqrt(r^2-(x(ind)-c)^2)
endfunction

function e=myfun(param,xm,ym)
e=demi_cercle(xm,param(1),param(2))-ym
endfunction
//c'est le tableau des différences,
xm (respectivement ym) est le tableau des abscisses (resp. ordonnées)
retourné par la fonction d'histogramme

function [f,r_opt,c_opt]=regression_demicerle(xm,ym,r0,c0)
[f,xopt]=leastsq(list(myfun,xm,ym),[r0;c0])
r_opt=xopt(1)
c_opt=xopt(2)
endfunction
// f est le maximum du carré de l'écart (carré de myfun)
entre la courbe théorique et les points expérimentaux

```

B.3 Temps de calcul

Scilab met environ n^3 secondes pour trouver les valeurs propres d'une matrice de taille $n \times 1000$.

C Densité jointe des valeurs propres des matrices GOE

En utilisant la définition des matrices GOE, on trouve facilement la densité de probabilité de ces matrices (on note $X = \text{vech}^{-1}(x)$) :

$$\begin{aligned}
f_M(x) &= \prod_{i=1}^n f_{i,i}(X_{i,i}) \prod_{i<j} f_{i,j}(X_{i,j}) \\
&= C \prod_{i=1}^n \exp\left(-\frac{(X_{i,i}-m)^2}{4\sigma^2}\right) \prod_{i<j} \exp\left(-\frac{X_{i,j}^2}{2\sigma^2}\right) \\
f_M(x) &= C \exp\left(-\frac{1}{4\sigma^2} \sum_{i,j} X_{i,j}^2 + \frac{m}{2\sigma^2} \sum_{i=1}^n X_{i,i} - \frac{mn}{4\sigma^2}\right)
\end{aligned}$$

En notant $\lambda_1, \dots, \lambda_n$ les valeurs propres (réelles) non classées de X on a donc :

$$f_M(x) = C \exp\left(-a \sum_{i=1}^n \lambda_i^2 + b \sum_{i=1}^n \lambda_i + c\right)$$

En faisant un changement de variables et en intégrant on trouve que la densité jointe des valeurs propres de $Y = \frac{1}{\sqrt{n}}X$ est (en prenant $\sigma^2 = 1$ et $m = 0$) :

$$g_n(\lambda_1, \dots, \lambda_n) = \tilde{C} \exp\left(-\frac{1}{2} \sum_{i=1}^n \lambda_i^2\right) \prod_{i < j} |\lambda_i - \lambda_j|$$

On reconnaît à droite la valeur absolue d'un déterminant de Vandermonde : ainsi, la densité jointe est invariante par permutation donc toutes les valeurs propres ont la même loi.

En outre, on obtient la densité à une valeur propre :

$$h_n(\lambda_1) = \tilde{C} \int \dots \int \left(-\frac{1}{2} \sum_{i=1}^n \lambda_i^2\right) \prod_{i < j} |\lambda_i - \lambda_j| d\lambda_2 \dots d\lambda_n$$

On remarquera que le paragraphe 1.2.2 affirme que :

$$\lim_{n \rightarrow +\infty} h_n\left(\frac{1}{2\sqrt{n}}\lambda\right) = \frac{2}{\pi} \sqrt{1 - \lambda^2} \mathbf{1}_{|\lambda| \leq 1}$$

D Une loi du χ^2

Soient X_1, \dots, X_n des variables gaussiennes centrées réduites indépendantes. On pose :

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

$$T^2 = \sum_{i=1}^n (X_i - M)^2 = \sum_{i=1}^n X_i^2 - nM^2$$

Théorème 10. *La variable T^2 suit une loi χ_{n-1}^2 .*

Démonstration. On va introduire les variables Y_i définies par $Y = HX$ où les X_i sont les coordonnées du vecteur X , idem pour Y et où H est une matrice orthogonale dont la dernière ligne est $(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$. Il faut remarquer qu'une telle matrice existe, par exemple :

$$\begin{pmatrix} \frac{-1}{\sqrt{2 \times 1}} & \frac{1}{\sqrt{2 \times 1}} & 0 & \dots & 0 \\ \frac{-1}{\sqrt{3 \times 2}} & \frac{-1}{\sqrt{3 \times 2}} & \frac{2}{\sqrt{3 \times 2}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{-1}{\sqrt{(n-1)(n-2)}} & \dots & \frac{-1}{\sqrt{(n-1)(n-1)}} & \frac{n-2}{\sqrt{(n-1)(n-2)}} & 0 \\ \frac{-1}{\sqrt{n(n-1)}} & \dots & \frac{-1}{\sqrt{n(n-1)}} & \frac{-1}{\sqrt{n(n-1)}} & \frac{n-1}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & \dots & \dots & \dots & \frac{1}{\sqrt{n}} \end{pmatrix}$$

qui est une matrice dite de Helmert. Cette matrice est orthogonale, donc $\|Y\|_2 = \|X\|_2$ ce qui s'écrit $Y_1^2 + \dots + Y_n^2 = X_1^2 + \dots + X_n^2$.

Les variables X_i étant des gaussiennes centrées réduites indépendantes, X est un vecteur gaussien de moyenne nulle et de matrice de covariance l'identité donc Y également (car H est orthogonale) : les Y_i sont donc des $\mathcal{N}(0, 1)$ indépendantes. En remarquant que $Y_n = \sqrt{n}M$, on a

$$\begin{aligned} T^2 &= X_1^2 + \cdots + X_{n-1}^2 + X_n^2 - nM^2 \\ &= Y_1^2 + \cdots + Y_{n-1}^2 + Y_n^2 - nM^2 \\ &= Y_1^2 + \cdots + Y_{n-1}^2 + nM^2 - nM^2 \\ T^2 &= Y_1^2 + \cdots + Y_{n-1}^2 \end{aligned}$$

La variable T^2 est une somme de carrés de $n - 1$ variables $\mathcal{N}(0, 1)$ indépendantes : T^2 suit donc une loi χ_{n-1}^2 . \square

E Termes d'ajouts non gaussiens

E.1 Uniforme discret

On suppose que les Y_{i_b} suivent une loi uniforme sur $\{-1, 1\}$; ils ont bien une espérance nulle et une variance de 1 (nécessaire pour que les coefficients de la matrice restent avec une moyenne nulle et une variance de 1). Les calculs menés jusqu'à présent restent valides, et on arrive donc à l'équation (10) :

$$\frac{\bar{\sigma}^2}{\sigma^2} = 1 - \rho + \rho(\overline{Y^2} - \bar{Y}^2)$$

On cherche encore une fois la loi de $\overline{Y^2} - \bar{Y}^2$. Le compte du premier terme est vite réglé, car comme Y prend les valeurs ± 1 on a $Y^2 = 1$ et donc $\overline{Y^2} = 1$. Pour trouver ce que vaut le deuxième terme, il suffit en fait de se rendre compte que $\frac{Y_{i_b} + 1}{2} \sim \mathcal{U}_{\{0,1\}}$. Ainsi,

$$\sum_{i_b=1}^m \frac{Y_{i_b} + 1}{2} \sim B$$

où $B \sim \mathcal{B}(m, \frac{1}{2})$ et on a finalement :

$$\bar{Y} \sim \frac{2B - m}{m} = \frac{2B}{m} - 1$$

Ainsi,

$$\frac{\bar{\sigma}^2}{\sigma^2} \sim 1 - \rho + \rho \left[1 - \left(\frac{2B}{m} - 1 \right)^2 \right] = 1 - \rho \left(\frac{2B}{m} - 1 \right)^2$$

On peut donc écrire en quelque sorte :

$$\mathbb{P} \left(\frac{\bar{\sigma}^2}{\sigma^2} = 1 - \rho \left(\frac{2k}{m} - 1 \right)^2 \right) = \binom{m}{k} \frac{1}{2^k} \frac{1}{2^{m-k}}$$

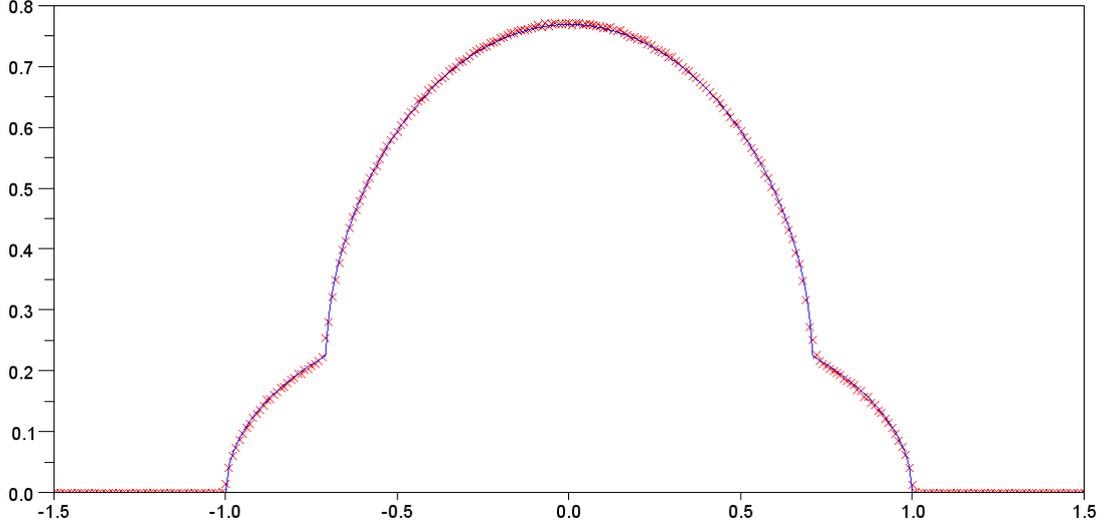


FIGURE 19 – Répartition des valeurs propres pour 3000 matrices de taille 1100 avec une corrélation à 2 blocs identiques et $\rho = 0.5$

et plus précisément :

$$\mathbb{P}_{\frac{\sigma^2}{\sigma^2}} = \frac{1}{2^m} \sum_{k=0}^m \binom{m}{k} \delta_{1-\rho\left(\frac{2k}{m}-1\right)^2}$$

et finalement :

$$\begin{aligned} D(x) &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u-x^2}}{u} \mathbb{P}_{\frac{\sigma^2}{\sigma^2}}(du) \\ &= \frac{2}{\pi} \sum_{k=0}^m \frac{1}{2^m} \binom{m}{k} \int_{x^2}^{+\infty} \frac{\sqrt{u-x^2}}{u} \delta_{1-\rho\left(\frac{2k}{m}-1\right)^2}(du) \\ D(x) &= \frac{2}{\pi} \sum_{k=0}^m \frac{1}{2^m} \binom{m}{k} \frac{\sqrt{1-\rho\left(\frac{2k}{m}-1\right)^2-x^2}}{1-\rho\left(\frac{2k}{m}-1\right)^2} \mathbf{1}_{x^2 \leq 1-\rho\left(\frac{2k}{m}-1\right)^2} \end{aligned}$$

En particulier, il y aura toujours un demi-cercle de « base » de rayon $\sqrt{1-\rho}$ (pour $k=0$) ; en revanche, seuls les cas m pair verront apparaître le demi-cercle de rayon 1 (pour $k = \frac{m}{2}$). Quelques simulations sont présentées dans les figures à 19 à 21 ; l’histogramme n’étant pas un demi-cercle, le phénomène n’est pas ergodique et on doit donc faire des moyennes d’ensemble. De plus, on doit prendre une taille de matrice assez grande pour que les demi-cercles soient bien formés ; les calculs sont donc très longs et les histogrammes ne sont pas parfaits (certains ont nécessité plus de 6h de calcul).

En fait, on peut également donner une expression de la densité pour des blocs de taille quelconque, plus précisément quand les rapports $t_{i_b} = \frac{\text{taille du bloc } i_b}{N_p}$ sont constants.

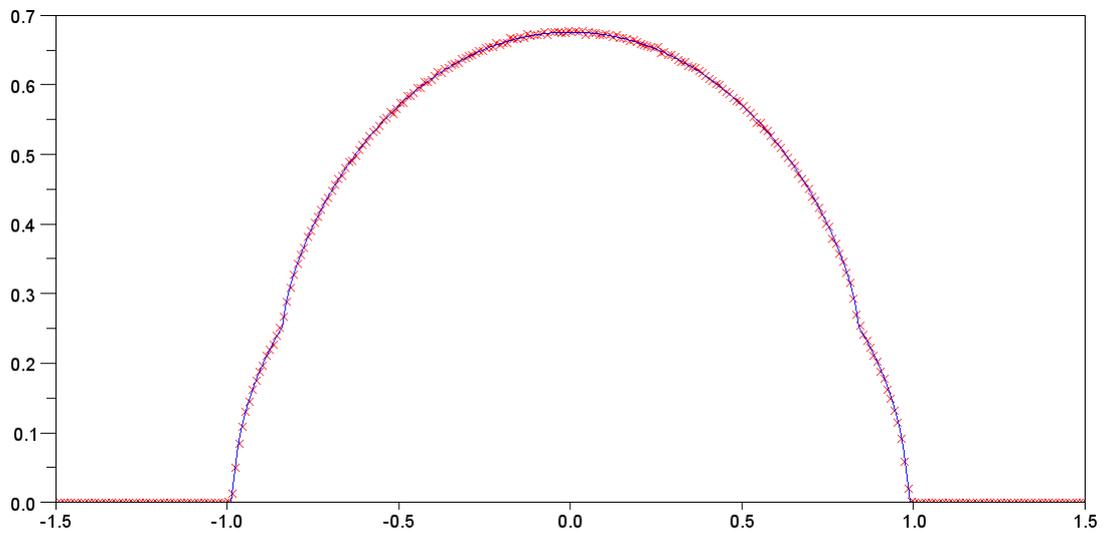


FIGURE 20 – Répartition des valeurs propres pour 3000 matrices de taille 1110 avec une corrélation à 3 blocs identiques et $\rho = 0.3$

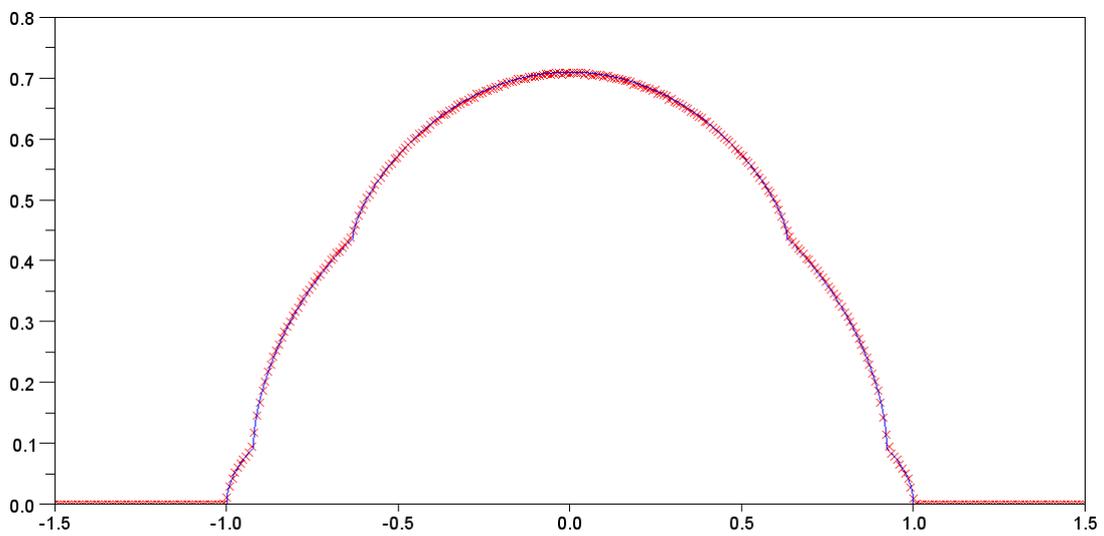


FIGURE 21 – Répartition des valeurs propres pour 6000 matrices de taille 1704 avec une corrélation à 4 blocs identiques et $\rho = 0.6$

Le terme

$$\overline{Y^2} - \bar{Y}^2 = \frac{1}{N_p} \sum_{i_b=1}^m \frac{N_p}{m} Y_{i_b}^2 - \left(\frac{1}{N_p} \sum_{i_b=1}^m \frac{N_p}{m} Y_{i_b} \right)^2$$

devient donc

$$\frac{1}{N_p} \sum_{i_b=1}^m N_p t_{i_b} Y_{i_b}^2 - \left(\frac{1}{N_p} \sum_{i_b=1}^m N_p t_{i_b} Y_{i_b} \right)^2 = \sum_{i_b=1}^m t_{i_b} Y_{i_b}^2 - \left(\sum_{i_b=1}^m t_{i_b} Y_{i_b} \right)^2$$

On a toujours $Y_{i_b}^2 = 1$, et comme $\sum t_{i_b} = 1$ il ne reste plus qu'à s'occuper du deuxième terme. Chaque Y_{i_b} à une probabilité $\frac{1}{2}$ de valoir -1 et une même probabilité de valoir 1 ; on a donc directement que, pour $(\varepsilon_1, \dots, \varepsilon_m) \in \{-1, 1\}^m$, la probabilité que $(Y_1, \dots, Y_m) = (\varepsilon_1, \dots, \varepsilon_m)$ est de $\frac{1}{2^m}$. On a $\frac{\sigma^2}{\sigma^2} = 1 - \rho(\sum_{i_b=1}^m t_{i_b} Y_{i_b})^2$, donc la loi de $\frac{\sigma^2}{\sigma^2}$ est :

$$\mathbb{P}_{\frac{\sigma^2}{\sigma^2}} = \sum_{(\varepsilon_1, \dots, \varepsilon_m) \in \{-1, 1\}^m} \frac{1}{2^m} \delta_{1 - \rho(\sum_{i_b=1}^m \varepsilon_{i_b} t_{i_b})^2}$$

ce que l'on peut écrire plus simplement, en notant $t = (t_1, \dots, t_m)$:

$$\mathbb{P}_{\frac{\sigma^2}{\sigma^2}} = \sum_{\varepsilon \in \{-1, 1\}^m} \frac{1}{2^m} \delta_{1 - \rho(\varepsilon \cdot t)^2}$$

Ainsi,

$$\begin{aligned} D(x) &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \mathbb{P}_{\frac{\sigma^2}{\sigma^2}}(du) \\ &= \frac{2}{\pi} \sum_{\varepsilon \in \{-1, 1\}^m} \frac{1}{2^m} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \delta_{1 - \rho(\varepsilon \cdot t)^2}(du) \\ D(x) &= \frac{2}{\pi} \sum_{\varepsilon \in \{-1, 1\}^m} \frac{1}{2^m} \frac{\sqrt{1 - \rho(\varepsilon \cdot t)^2 - x^2}}{1 - \rho(\varepsilon \cdot t)^2} \mathbf{1}_{x^2 \leq 1 - \rho(\varepsilon \cdot t)^2} \end{aligned}$$

On peut voir ce que cela donne avec des blocs de rapports $[\frac{1}{2}, \frac{1}{4}, \frac{1}{4}]$. Les huit combinaisons possibles sont :

$$\begin{array}{ll} +\frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1 & -\frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 0 \\ +\frac{1}{2} + \frac{1}{4} - \frac{1}{4} = \frac{1}{2} & -\frac{1}{2} + \frac{1}{4} - \frac{1}{4} = -\frac{1}{2} \\ +\frac{1}{2} - \frac{1}{4} + \frac{1}{4} = \frac{1}{2} & -\frac{1}{2} - \frac{1}{4} + \frac{1}{4} = -\frac{1}{2} \\ +\frac{1}{2} - \frac{1}{4} - \frac{1}{4} = 0 & -\frac{1}{2} - \frac{1}{4} - \frac{1}{4} = -1 \end{array}$$

On verra donc apparaître trois demi-cercles, dont les couples (carré du rayon, occurrences) sont $(1 - \rho, 2)$, $(1 - \frac{\rho}{4}, 4)$ et $(1, 2)$. Ainsi, la densité des valeurs propres est donnée par :

$$D(x) = \frac{2}{\pi} \frac{1}{2^3} \left[2 \frac{\sqrt{1 - \rho - x^2}}{1 - \rho} \mathbf{1}_{x^2 \leq 1 - \rho} + 4 \frac{\sqrt{1 - \frac{\rho}{4} - x^2}}{1 - \frac{\rho}{4}} \mathbf{1}_{x^2 \leq 1 - \frac{\rho}{4}} + 2 \sqrt{1 - x^2} \mathbf{1}_{x^2 \leq 1} \right]$$

On peut voir sur la figure 22 le résultat.

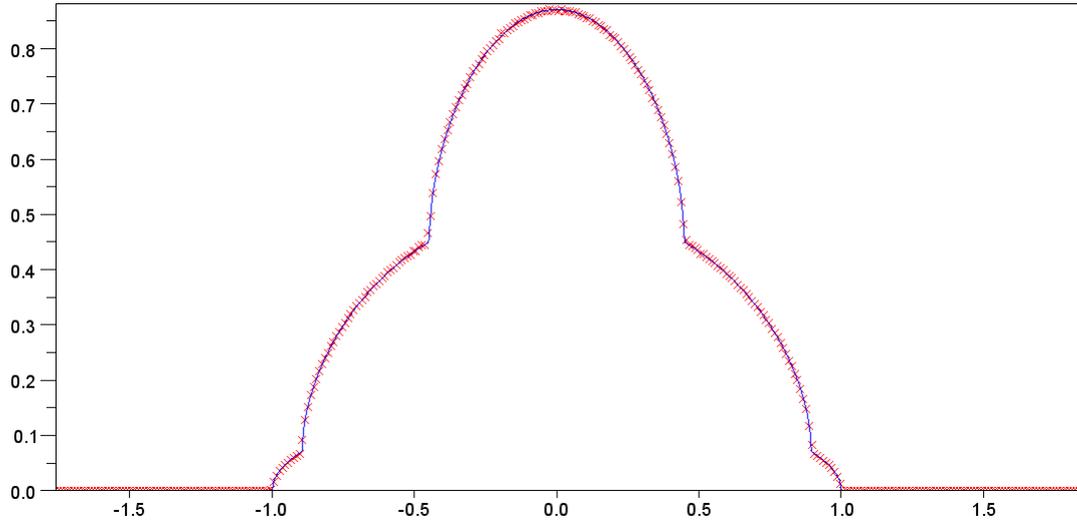


FIGURE 22 – Répartition des valeurs propres pour 1000 matrices de taille 3000 avec une structure de corrélation de blocs $[\frac{1}{2}, \frac{1}{4}, \frac{1}{4}]$ et $\rho = 0.8$

E.2 Uniforme continu

La variance d'une loi uniforme sur $[-1, 1]$ est $\int_{-1}^1 x^2 dx - 0^2 = \frac{2}{3}$: pour obtenir une variance de 1, on va donc choisir $Y_{b(i)} \sim \sqrt{\frac{3}{2}}\mathcal{U}_{[-1,1]}$. Les calculs menés restent toujours valides, et on a donc à l'équation (10) :

$$\frac{\bar{\sigma}^2}{\sigma^2} = 1 - \rho + \rho(\overline{Y^2} - \bar{Y}^2)$$

Reste donc à trouver la loi de $\overline{Y^2} - \bar{Y}^2$; or voilà, c'est assez compliqué, et donc pour simplifier on va se restreindre au cas où $m = 2$. On a donc :

$$\begin{aligned} \overline{Y^2} - \bar{Y}^2 &= \frac{1}{2}(Y_1^2 + Y_2^2) - \left[\frac{1}{2}(Y_1 + Y_2)\right]^2 \\ &= \frac{1}{2}(Y_1^2 + Y_2^2) - \frac{1}{4}(Y_1^2 + Y_2^2) - \frac{1}{2}Y_1Y_2 \\ &= \frac{1}{4}(Y_1^2 + Y_2^2 - 2Y_1Y_2) \\ &= \frac{1}{4}(Y_1 - Y_2)^2 \\ \overline{Y^2} - \bar{Y}^2 &= \frac{3}{8}(U_1 - U_2)^2 \end{aligned}$$

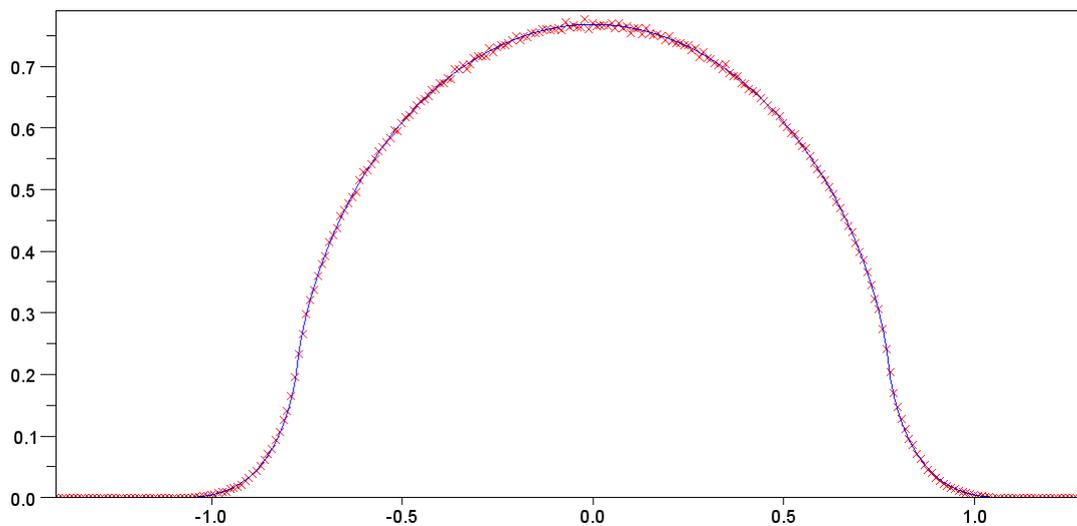


FIGURE 23 – Répartition des valeurs propres pour 1000 matrices de taille 1000 avec 2 blocs identiques et $\rho = 0.4$

où $Y_i = \sqrt{\frac{3}{2}}U_i$, U_i suivant la loi uniforme sur $[-1, 1]$. À l'aide d'un produit de convolution, on trouve que la densité de $U_1 - U_2$ est $x \mapsto \frac{2-|x|}{4}\mathbf{1}_{|x|\leq 2}$ puis que celle de $\frac{\sigma^2}{\sigma^2}$ est :

$$u \mapsto \frac{2}{3\rho} \left[\sqrt{\frac{3\rho}{2(u - (1 - \rho))}} - 1 \right] \mathbf{1}_{1-\rho < u \leq 1 + \frac{\rho}{2}}$$

La densité des valeurs propres devient :

$$\begin{aligned} D(x) &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \mathbb{P}_{\frac{\sigma^2}{\sigma^2}}(du) \\ &= \frac{2}{\pi} \int_{x^2}^{+\infty} \frac{\sqrt{u - x^2}}{u} \frac{2}{3\rho} \left[\sqrt{\frac{3\rho}{2(u - (1 - \rho))}} - 1 \right] \mathbf{1}_{1-\rho < u \leq 1 + \frac{\rho}{2}} du \\ D(x) &= \frac{4}{3\pi\rho} \int_{\max(x^2, 1-\rho)}^{\max(x^2, 1 + \frac{\rho}{2})} \frac{\sqrt{u - x^2}}{u} \left[\sqrt{\frac{3\rho}{2(u - (1 - \rho))}} - 1 \right] du \end{aligned}$$

En particulier, on remarquera que cette densité est bornée; la figure 23 donne un exemple d'histogramme.