

# Devoir surveillé n°7

MP Clemenceau 2022-23

Jeudi 16 mars 2023

Vous avez 4 heures dans la joie et la bonne humeur mais en silence !!

Il n'est proposé qu'un seul sujet : CentraleSupélec 2021

Il sera tenu compte de la présentation et de la rigueur des démonstrations. Toute copie non rédigée ne sera pas corrigée. Il est demandé aux étudiants de mettre leurs nom et prénom sur chaque copie et de numéroter ces dites copies.

**Les calculatrices sont interdites.**

**Lorsqu'un raisonnement utilise le résultat d'une question précédente, il est demandé au candidat d'indiquer précisément le numéro de la question utilisée.**



## *La loi du demi-cercle*

### Notations

- Pour tous entiers naturels  $p$  et  $q$  tels que  $p \leq q$  on note  $\llbracket p, q \rrbracket$  l'ensemble  $\{i \in \mathbb{N} \mid p \leq i \leq q\}$ .
- Pour tout entier naturel  $n$  tel que  $n \geq 1$ , on note  $\mathcal{M}_n(\mathbb{R})$  l'ensemble des matrices à  $n$  lignes et  $n$  colonnes à coefficients réels,  $\mathcal{S}_n(\mathbb{R})$  l'ensemble des matrices symétriques et à coefficients réels de taille  $n$  et  $\mathcal{O}_n(\mathbb{R})$  l'ensemble des matrices réelles orthogonales de taille  $n$ .
- Pour toute matrice symétrique réelle  $M \in \mathcal{S}_n(\mathbb{R})$  on note  $\lambda_1(M) \geq \lambda_2(M) \geq \dots \geq \lambda_n(M)$  ses valeurs propres rangées dans l'ordre décroissant.
- On note  $M^T$  la transposée d'une matrice  $M$ .
- Pour tout  $(i, j) \in \llbracket 1, n \rrbracket^2$ , on note  $E_{ij}$  la matrice de la base canonique de  $\mathcal{M}_n(\mathbb{R})$  dont tous les coefficients sont nuls sauf celui situé sur la  $i$ -ème ligne et la  $j$ -ième colonne qui vaut 1 .
- Pour toute matrice  $M \in \mathcal{M}_n(\mathbb{R})$ , on note  $\|M\|_F = \sqrt{\text{tr}(MM^T)}$  sa norme euclidienne canonique.
- Dans tout le problème, on note  $(\Omega, \mathcal{A}, \mathbf{P})$  un espace probabilisé. Toutes les variables aléatoires considérées sont définies sur  $\Omega$ .
- Étant donnée une variable aléatoire discrète  $X$  à valeurs réelles admettant une espérance, on note  $\mathbf{E}(X)$  son espérance. Si  $X$  admet une variance, on note  $\mathbf{V}(X)$  sa variance.

### Problématique

En essayant d'expliquer la répartition des niveaux d'énergie des noyaux des atomes lourds, Eugène Wigner a été amené, dans les années 1950, à étudier le spectre de matrices symétriques réelles aléatoires de grande taille. La figure 1 montre un histogramme qui représente la répartition des valeurs propres d'une matrice symétrique réelle de taille  $n = 2500$  constituée de variables aléatoires indépendantes identiquement distribuées, d'espérance nulle et de variance égale à 1.

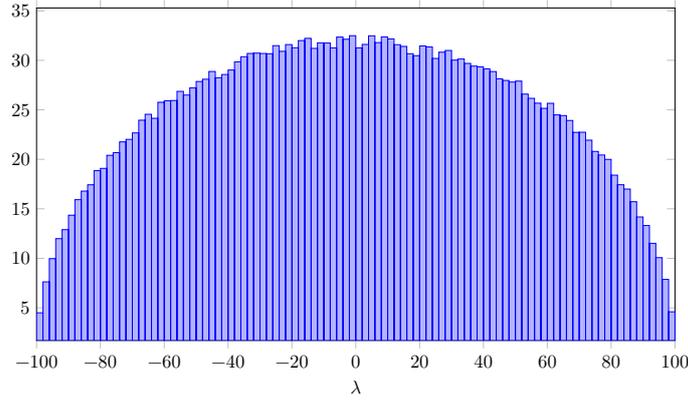


Figure 1

On voit que les valeurs propres se répartissent suivant un profil en demi-cercle de rayon  $2\sqrt{n}$ . En normalisant par un facteur  $1/\sqrt{n}$  on obtient le profil d'un demi-cercle de rayon 2.

Plus précisément, on considère  $(X_{ij})_{(i,j) \in (\mathbb{N}^*)^2}$  une famille de variables aléatoires discrètes réelles telles que

- pour tout  $(i, j) \in (\mathbb{N}^*)^2$ ,  $X_{ij} = X_{ji}$  ;
- les variables aléatoires  $X_{ij}$  sont de même loi, d'espérance nulle et de variance 1 ;
- pour  $n \geq 1$ , les variables aléatoires  $X_{ij}$ , pour  $1 \leq i \leq j \leq n$ , sont mutuellement indépendantes.

Pour tout  $n \in \mathbb{N}^*$  et pour tout  $\omega \in \Omega$ , on note  $M_n(\omega)$  la matrice  $(X_{ij}(\omega))_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{R})$ .

Pour tout  $\omega \in \Omega$ , on note  $\Lambda_{1,n}(\omega) \geq \dots \geq \Lambda_{n,n}(\omega)$  les valeurs propres de  $\frac{1}{\sqrt{n}} M_n(\omega)$ .

On définit ainsi, pour  $n \in \mathbb{N}^*$  et  $i \in \llbracket 1, n \rrbracket$ , des variables aléatoires réelles discrètes  $\Lambda_{i,n}$  sur  $(\Omega, \mathcal{A}, \mathbf{P})$ .

On se propose de montrer le résultat suivant :

**Loi du demi-cercle**

Pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$ , continue et bornée,

$$\mathbf{E} \left( \frac{1}{n} \sum_{i=1}^n f(\Lambda_{i,n}) \right) \xrightarrow{n \rightarrow +\infty} \frac{1}{2\pi} \int_{-2}^2 f(x) \sqrt{4-x^2} \, dx$$

Dans la partie I, on établit une inégalité portant sur les valeurs propres d'un couple de matrices symétriques. La partie II est consacrée à la résolution d'un problème de dénombrement qui est utilisée dans la partie III où la loi du demi-cercle est démontrée pour des variables aléatoires uniformément bornées. Dans la partie IV, on établit la loi du demi-cercle dans le cas général en utilisant les résultats des parties I et III.

## I Inégalité de Hoffman-Wielandt

**I.A** - Soient  $A$  et  $B$  deux matrices de  $\mathcal{S}_n(\mathbb{R})$ .

**Q 1)** Montrer que, pour tout  $M$  dans  $\mathcal{M}_n(\mathbb{R})$  et pour tous  $P$  et  $Q$  dans  $\mathcal{O}_n(\mathbb{R})$ , on a  $\|PMQ\|_F = \|M\|_F$ .

**Q 2)** On note  $D_A = \text{diag}(\lambda_1(A), \dots, \lambda_n(A))$  et  $D_B = \text{diag}(\lambda_1(B), \dots, \lambda_n(B))$ . Montrer qu'il existe une matrice orthogonale  $P = (p_{i,j})_{1 \leq i, j \leq n}$  telle que  $\|A - B\|_F^2 = \|D_A P - P D_B\|_F^2$

**Q 3)** Montrer que

$$\|A - B\|_F^2 = \sum_{1 \leq i, j \leq n} p_{i,j}^2 (\lambda_i(A) - \lambda_j(B))^2$$

**I.B** - On note  $\mathcal{B}_n(\mathbb{R})$  l'ensemble des matrices bistochastiques de  $\mathcal{M}_n(\mathbb{R})$ , c'est-à-dire l'ensemble des matrices  $M = (m_{i,j})_{1 \leq i, j \leq n}$  dont tous les coefficients sont positifs ou nuls et tels que  $\sum_{j=1}^n m_{i,j} = \sum_{j=1}^n m_{j,i} = 1$  pour tout  $i \in \llbracket 1, n \rrbracket$ .

$$\text{On note } f : \begin{cases} \mathcal{M}_n(\mathbb{R}) & \rightarrow & \mathbb{R} \\ M & \mapsto & \sum_{1 \leq i, j \leq n} m_{i,j} (\lambda_i(A) - \lambda_j(B))^2 \end{cases} .$$

**Q 4)** Justifier que  $f$  admet un minimum sur  $\mathcal{B}_n(\mathbb{R})$ .

On se propose de montrer que ce minimum est atteint en la matrice identité.

**Q 5)** Soit  $(i, j, k) \in \llbracket 1, n \rrbracket^3$  tel que  $j \geq i$  et  $k \geq i$ . Montrer que, pour  $M \in \mathcal{M}_n(\mathbb{R})$  et pour  $x \in \mathbb{R}^+$ ,

$$f(M + xE_{ii} + xE_{jk} - xE_{ik} - xE_{ji}) - f(M) = 2x(\lambda_i(A) - \lambda_j(A))(\lambda_k(B) - \lambda_i(B)) \leq 0$$

**Q 6)** Soient  $n \geq 2$  et  $M = (m_{i,j})_{1 \leq i, j \leq n} \in \mathcal{B}_n(\mathbb{R})$  une matrice différente de l'identité. On note  $i$  le plus petit entier appartenant à  $\llbracket 1, n \rrbracket$  tel que  $m_{i,i} \neq 1$ . Montrer qu'il existe une matrice  $M' = (m'_{i,j})_{1 \leq i, j \leq n} \in \mathcal{B}_n(\mathbb{R})$  telle que  $f(M') \leq f(M)$  et  $m'_{j,j} = 1$  pour tout  $j \in \llbracket 1, i \rrbracket$ .

**Q 7)** En déduire que

$$\min \{f(M) \mid M \in \mathcal{B}_n(\mathbb{R})\} = f(I_n)$$

**I.C -**

**Q 8)** En déduire que

$$\forall (A, B) \in S_n(\mathbb{R})^2, \quad \sum_{i=1}^n (\lambda_i(A) - \lambda_i(B))^2 \leq \|A - B\|_F^2$$

## II Dénombrement des mots bien parenthésés

Dans cette partie, on s'intéresse à des chaînes de caractères constituées uniquement des deux caractères parenthèse ouvrante et parenthèse fermante. On dit qu'un mot est *bien parenthésé* s'il commence par une parenthèse ouvrante et qu'à toute parenthèse ouvrante est associée une (unique) parenthèse fermante *qui lui est postérieure*. Par exemple le mot

( ) ( ( ) )

est bien parenthésé. En revanche, le mot

( ) ) (

n'est pas bien parenthésé. Un mot bien parenthésé est ainsi forcément constitué d'un nombre pair de caractères, chaque parenthèse qui s'ouvre doit se refermer.

Pour tout entier  $n \geq 1$ , on note  $C_n$  le nombre de mots bien parenthésés de longueur  $2n$ . On pose par commodité  $C_0 = 1$ .

**II.A -**

**Q 9)** En énumérant les différents mots bien parenthésés de longueur 2, 4 et 6, montrer que  $C_1 = 1, C_2 = 2$  et déterminer  $C_3$ .

**Q 10)** Montrer que, pour tout entier naturel  $n, C_n \leq 2^{2n}$ . Que peut-on en déduire pour le rayon de convergence de la série entière  $\sum C_k x^k$  ?

**Q 11)** Montrer par un raisonnement combinatoire que, pour tout entier  $k \geq 1$ ,

$$C_k = \sum_{i=0}^{k-1} C_i C_{k-i-1}$$

On peut remarquer qu'un mot bien parenthésé est forcément de la forme  $(m)m'$  avec  $m$  et  $m'$  deux mots bien parenthésés, éventuellement vides.

**II.B -** Pour tout  $x \in ]-\frac{1}{4}, \frac{1}{4}[$ , on pose  $F(x) = \sum_{k=0}^{+\infty} C_k x^k$ .

**Q 12)** Montrer que, pour tout  $x \in ]-\frac{1}{4}, \frac{1}{4}[$ ,  $F(x) = 1 + x(F(x))^2$ .

**Q 13)** Montrer que la fonction  $f : \begin{cases} ]-\frac{1}{4}, \frac{1}{4}[ & \rightarrow & \mathbb{R} \\ x & \mapsto & 2xF(x) - 1 \end{cases}$  ne s'annule pas.

**Q 14)** Déterminer, pour tout  $x \in ]-\frac{1}{4}, \frac{1}{4}[$ , une expression de  $F(x)$  en fonction de  $x$ .

**Q 15)** Déterminer le développement en série entière de la fonction  $u \mapsto \sqrt{1-u}$ . On écrira les coefficients sous la forme d'un quotient de factorielles et de puissances de 2.

**Q 16)** Montrer que, pour tout entier naturel  $n$ ,

$$C_n = \frac{(2n)!}{(n+1)!n!}$$

### III Loi du demi-cercle, cas uniformément borné

On suppose, uniquement dans cette partie, que les variables aléatoires  $X_{ij}$  sont uniformément bornées :

$$\exists K \in \mathbb{R}, \forall (i, j) \in (\mathbb{N}^*)^2, |X_{ij}| \leq K.$$

**III.A** - Pour tout entier naturel  $k$  on pose

$$m_k = \frac{1}{2\pi} \int_{-2}^2 x^k \sqrt{4-x^2} dx$$

**Q 17)** Pour  $k \in \mathbb{N}$ , que vaut  $m_{2k+1}$  ?

**Q 18)** En utilisant le changement de variable  $x = 2 \sin t$ , calculer  $m_0$ .

**Q 19)** À l'aide d'une intégration par parties, montrer que, pour tout entier naturel  $k$ ,

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

**Q 20)** En déduire que

$$m_k = \begin{cases} C_{k/2} & \text{si } k \text{ est pair,} \\ 0 & \text{si } k \text{ est impair.} \end{cases}$$

**III.B** - Soit  $k$  un entier naturel. On se propose de montrer que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n \Lambda_{i,n}^k \right) = m_k,$$

où  $\Lambda_{i,n}^k = (\Lambda_{i,n})^k$  est la puissance  $k$ -ième de  $\Lambda_{i,n}$ .

**Q 21)** Justifier que la variable aléatoire  $\sum_{i=1}^n \Lambda_{i,n}^k$  admet une espérance et que

$$\mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n \Lambda_{i,n}^k \right) = \frac{1}{n^{1+k/2}} \mathbb{E} (\text{tr} (M_n^k)) = \frac{1}{n^{1+k/2}} \sum_{(i_1, \dots, i_k) \in \llbracket 1, n \rrbracket^k} \mathbb{E} (X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}).$$

On appelle *cycle* de longueur  $k$  à valeurs dans  $\llbracket 1, n \rrbracket$ , tout  $(k+1)$ -uplet  $\vec{i} = (i_1, i_2, \dots, i_k, i_1)$  d'éléments de  $\llbracket 1, n \rrbracket$ . Les éléments  $i_1, \dots, i_k$  sont appelés *sommets* du cycle  $\vec{i}$ . On dit aussi que le cycle *passse* par ces sommets. On note  $|\vec{i}|$  le nombre de sommets distincts du cycle  $\vec{i}$ .

On appelle *arêtes* du cycle  $(i_1, i_2, \dots, i_k, i_1)$  les couples non ordonnés (l'ordre des deux éléments de chaque couple n'est pas significatif)  $(i_1, i_2), (i_2, i_3), \dots, (i_k, i_1)$ .

Par exemple  $\vec{i} = (1, 3, 5, 3, 2, 1)$  est un cycle de longueur 6 dans  $\llbracket 1, 5 \rrbracket$ . Les sommets de ce cycle sont les éléments 1, 2, 3 et 5, donc  $|\vec{i}| = 4$ . Les arêtes distinctes de ce cycle sont  $(1, 3), (3, 5), (3, 2), (2, 2)$  et  $(2, 1)$ . Les arêtes  $(3, 5)$  et  $(5, 3)$  sont les mêmes.

**Q 22)** Montrer que le nombre de cycles de longueur  $k$  dans  $\llbracket 1, n \rrbracket$  passant par  $\ell$  sommets distincts est inférieur ou égal à  $n^\ell \ell^k$ .

**Q 23)** En déduire que

$$\frac{1}{n^{1+k/2}} \sum_{\substack{\vec{i} \in \llbracket 1, n \rrbracket^k \\ |\vec{i}| \leq (k+1)/2}} |\mathbb{E} (X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1})| \xrightarrow{n \rightarrow +\infty} 0$$

On classe les cycles de longueur  $k$  en trois sous-ensembles :

- l'ensemble  $\mathcal{A}_k$ , constitué des cycles où au moins une arête n'apparaît qu'une fois ;
- l'ensemble  $\mathcal{B}_k$ , constitué des cycles où toutes les arêtes apparaissent exactement deux fois ;
- l'ensemble  $\mathcal{C}_k$ , constitué des cycles où toutes les arêtes apparaissent au moins deux fois et il en existe au moins une qui apparaît au moins trois fois.

**Q 24)** Montrer que, si le cycle  $(i_1, i_2, \dots, i_k, i_1)$  appartient à  $\mathcal{A}_k$ , alors

$$\mathbb{E} (X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}) = 0$$

**Q 25)** Montrer que, pour tout cycle  $\vec{\nu}$  appartenant à  $\mathcal{C}_k$ ,  $|\vec{\nu}| \leq \frac{k+1}{2}$ .

**Q 26)** Que peut-on dire de  $\mathcal{B}_k$  si  $k$  est impair ? En déduire que  $\lim_{n \rightarrow +\infty} \mathbf{E} \left( \frac{1}{n} \sum_{i=1}^n \Lambda_{i,n}^k \right) = 0$  dans ce cas.

On suppose dans la suite que  $k$  est pair et que  $\vec{\nu} \in \mathcal{B}_k$  est un cycle passant par  $\frac{k}{2} + 1$  sommets distincts. Autrement dit  $|\vec{\nu}| = \frac{k}{2} + 1$ .

On parcourt les arêtes de  $\vec{\nu}$  dans l'ordre. À chaque arête de  $\vec{\nu}$  on associe une parenthèse ouvrante si cette arête apparaît pour la première fois et une parenthèse fermante si elle apparaît pour la deuxième fois. Par exemple, au cycle  $(1, 3, 2, 3, 1)$  correspond  $(( ))$ , au cycle  $(1, 2, 1, 3, 1)$  correspond  $(( ))$ .

**Q 27)** Justifier que l'on obtient ainsi un mot bien parenthésé de longueur  $k$ .

**Q 28)** Dénombrer les cycles  $\vec{\nu}$  qui correspondent à un mot bien parenthésé fixé.

**Q 29)** Déduire de ce qui précède que

$$\lim_{n \rightarrow +\infty} \mathbf{E} \left( \frac{1}{n} \sum_{i=1}^n \Lambda_{i,n}^k \right) = C_{k/2}$$

### III.C -

**Q 30)** En déduire que, pour tout polynôme  $P \in \mathbb{R}[X]$ ,

$$\lim_{n \rightarrow +\infty} \mathbf{E} \left( \frac{1}{n} \sum_{i=1}^n P(\Lambda_{i,n}) \right) = \frac{1}{2\pi} \int_{-2}^2 P(x) \sqrt{4-x^2} dx$$

### III.D - Soit $A > 2$ .

**Q 31)** Montrer que, pour tout  $(p, q) \in \mathbb{N}^2$ ,

$$\mathbf{E} \left( \sum_{\substack{1 \leq i \leq n \\ |\Lambda_{i,n}| \geq A}} |\Lambda_{i,n}|^p \right) \leq \frac{1}{A^{p+2q}} \mathbf{E} \left( \sum_{i=1}^n |\Lambda_{i,n}|^{2(p+q)} \right)$$

**Q 32)** Montrer que

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathbf{E} \left( \sum_{\substack{1 \leq i \leq n \\ |\Lambda_{i,n}| \geq A}} |\Lambda_{i,n}|^p \right) = 0$$

**Q 33)** Soient  $f$  une fonction continue et bornée de  $\mathbb{R}$  dans  $\mathbb{R}$  et  $P$  un polynôme de degré  $p$ . Justifier qu'il existe une constante  $K$  telle que

$$\forall x \in \mathbb{R} \setminus ]-A, A[, |f(x) - P(x)| \leq K|x|^p$$

**Q 34)** En déduire que

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathbf{E} \left( \sum_{\substack{1 \leq i \leq n \\ |\Lambda_{i,n}| \geq A}} |f - P|(\Lambda_{i,n}) \right) = 0$$

### III.E -

**Q 35)** Soit  $f$  une fonction continue et bornée de  $\mathbb{R}$  dans  $\mathbb{R}$ . Montrer que

$$\lim_{n \rightarrow +\infty} \mathbf{E} \left( \frac{1}{n} \sum_{i=1}^n f(\Lambda_{i,n}) \right) = \frac{1}{2\pi} \int_{-2}^2 f(x) \sqrt{4-x^2} dx$$

## IV Loi du demi-cercle, cas général

On revient au cas général. On note  $\mathbb{1}_A$  la variable aléatoire indicatrice d'un événement  $A$ .

Pour tout  $(i, j) \in (\mathbb{N}^*)^2$  et pour tout  $C > 0$ , on pose

$$\sigma_{ij}(C) = \sqrt{\mathbf{V}(X_{ij} \mathbb{1}_{|X_{ij}| \leq C})}$$

Si  $\sigma_{ij}(C) \neq 0$ , on pose

$$\widehat{X}_{ij}(C) = \frac{1}{\sigma_{ij}(C)} (X_{ij} \mathbb{1}_{|X_{ij}| \leq C} - \mathbf{E}(X_{ij} \mathbb{1}_{|X_{ij}| \leq C}))$$

### IV.A -

**Q 36)** Soit  $X$  une variable aléatoire discrète d'espérance finie. Montrer que

$$\mathbf{E}(X\mathbf{1}_{|X|\leq C}) \xrightarrow{C\rightarrow+\infty} \mathbf{E}(X)$$

**Q 37)** En déduire que

$$\lim_{C\rightarrow+\infty} \sigma_{ij}(C) = 1$$

**Q 38)** Justifier que, pour  $C$  assez grand, les variables  $\widehat{X}_{ij}(C)$  sont bien définies et qu'elles sont alors bornées, centrées, de variance 1 et qu'elles sont mutuellement indépendantes pour  $1 \leq i \leq j$ .

**Q 39)** Montrer que

$$X_{ij} - \widehat{X}_{ij}(C) = \left(1 - \frac{1}{\sigma_{ij}(C)}\right) X_{ij} + \frac{1}{\sigma_{ij}(C)} (X_{ij}\mathbf{1}_{|X_{ij}|>C} - \mathbf{E}(X_{ij}\mathbf{1}_{|X_{ij}|>C}))$$

**Q 40)** Montrer que

$$\lim_{C\rightarrow+\infty} \mathbf{E}\left(\left(X_{ij} - \widehat{X}_{ij}(C)\right)^2\right) = 0$$

**IV.B -** Pour tout entier  $n$  tel que  $n \geq 1$ , on note  $\widehat{M}_n(C) = \left(\widehat{X}_{ij}(C)\right)_{1 \leq i, j \leq n}$ . Pour tout  $\omega \in \Omega$ , on note  $\widehat{\Lambda}_{1,n}(\omega) \geq \dots \geq \widehat{\Lambda}_{n,n}(\omega)$  les valeurs propres de  $\frac{1}{\sqrt{n}}\widehat{M}_n(C)(\omega)$  rangées dans l'ordre décroissant.

On obtient ainsi des variables aléatoires réelles discrètes  $\widehat{\Lambda}_{1,n}, \dots, \widehat{\Lambda}_{n,n}$  sur  $(\Omega, \mathcal{A}, \mathbf{P})$ .

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction  $K$ -lipschitzienne.

**Q 41)** Montrer que

$$\left| \mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n f(\Lambda_{i,n})\right) - \mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n f(\widehat{\Lambda}_{i,n})\right) \right| \leq \frac{K}{n} \mathbf{E}\left(\|M_n - \widehat{M}_n(C)\|_F\right)$$

**Q 42)** On suppose de plus  $f$  bornée. Montrer

$$\mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n f(\Lambda_{i,n})\right) \xrightarrow{n\rightarrow+\infty} \frac{1}{2\pi} \int_{-2}^2 f(x) \sqrt{4-x^2} dx$$

**IV.C -**

**Q 43)** Montrer la loi du demi-cercle dans le cas général.