Introduction au calcul des matrices aléatoires

Jamal Najim
CNRS et ENST
najim@enst.fr

Journée du GDR ISIS,
le 10 janvier 2007.

Présentations bientôt disponibles en ligne
http://www.tsi.enst.fr/∼najim/gdr/

Introduction

On considère une matrice Y_n dont les entrées (réelles ou complexes) sont des variables aléatoires dont il faudra préciser la nature. La matrice Y_n est de taille $N \times n$ où:

$$\frac{N}{n} \xrightarrow[n\to\infty]{} c \in (0, \infty).$$

On s'intéresse à la matrice de Gram YY^T (ou YY^* si Y est complexe) et surtout à son spectre lorsque n tend vers ∞.

Le modèle aléatoire pour Y

On s'intéressera au modèle suivant:

$$Y = \frac{1}{\sqrt{n}} D^{\frac{1}{2}} X \tilde{D}^{\frac{1}{2}} + A$$

où:

- X est une matrice aléatoire de taille $N \times n$ dont les entrées X_{ij} sont centrées, **indépendantes et identiquement distribuées** (la loi ne dépend pas de n);
- A est une matrice déterministe de taille $N \times n$. A est appelée **matrice de centrage**.
- D est une matrice diagonale de taille $N \times N$ dont les éléments sont positifs. D est appelée **matrice de corrélation à gauche**;
- $\tilde{D}$ est une matrice diagonale de taille $n \times n$ dont les éléments sont positifs. $\tilde{D}$ est appelée **matrice de corrélation à droite**;

Quelques cas plus simples

Dans la suite, on distinguera les modèles simplifiés suivants:

1. le **modèle i.i.d. centré**:
$$Y = \frac{1}{\sqrt{n}} X;$$
2. Le **modèle i.i.d. non centré**:
$$Y = \frac{1}{\sqrt{n}} X + A;$$
3. Le **modèle à profil de variance séparable non centré**:
$$Y = \frac{1}{\sqrt{n}} D^{\frac{1}{2}} X \tilde{D}^{\frac{1}{2}} + A.$$

Etude du spectre de YY^T

La matrice YY^T est une matrice diagonalisable. Ses valeurs propres

$$(\lambda_i, 1 \leq i \leq N)$$

sont positives. Elles dépendent des entrées de Y est sont à ce titre **aléatoires**.

On appelle mesure spectrale associée à YY^T la mesure de probabilité **aléatoire**

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}(d\lambda),$$

autrement dit, si A est un intervalle de $\mathbb{R}^+$, alors

$$L_N(A) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}(A) = \frac{\#\{\lambda_i \in A\}}{N}.$$

- Notre premier objectif est de comprendre le comportement de L_N quand $n \to \infty$ et $\frac{N}{n} \to c \in (0, \infty)$.

Etude du **logdet**

Considérons la fonctionnelle suivante:

$$I(\rho) = \frac{1}{N} \mathbb{E} \log \det \left(I_N + \frac{Y_n Y_n^T}{\rho} \right).$$

Cette fonctionnelle est très importante en communications numériques (cf. supra - exposé de W. Hachem). Elle s'exprime en fonction du spectre:

$$I(\rho) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \log \left(1 + \frac{\lambda_i}{\rho} \right)$$

Nos objectifs sont

- d'approximer cette fonctionnelle par une quantité qui se calcule facilement,
- éventuellement de comprendre les fluctuations de la quantité sous l' espérance:

$$\frac{1}{N} \log \det \left(I_N + \frac{Y_n Y_n^T}{\rho} \right) = \frac{1}{N} \sum_{i=1}^{N} \log \left(1 + \frac{\lambda_i}{\rho} \right).$$

Définition formelle

Comme la fonction caractéristiques ou la fonction de répartition, la transformée de Stieltjes est une fonction qui permet de manipuler les mesures de probabilités (ou plus généralement les mesures positives) **sans perte d'information**. Formellement, on la définit par:

$$\mathbf{f}(z) = \int_{\mathbb{R}^+} \frac{\mu(d\lambda)}{\lambda - z}$$

où μ est une mesure positive définie sur $\mathbb{R}^+$ et $z \in \mathbb{C} \setminus \mathbb{R}^+$ est un nombre complexe.

Quelques propriétés

La formule d'inversion Si je connais la mesure μ, je connais $\mathbf{f}$, et réciproquement:

$$\mu(a,b) = \lim_{y \to 0} \frac{1}{\pi} \int_a^b \mathrm{Im}\mathbf{f}(x+\mathbf{i}y)dx.$$

Autrement dit, $\mathbf{f}$ me permet de caractériser complètement μ.

Caractérisation d'une transformée de Stieltjes Réciproquement, si $\mathbf{f}$ est une fonction de $\mathbb{C} \setminus \mathbb{R}^+$ dans $\mathbb{C}$ qui satisfait un certain nombre de conditions techniques, alors il existe une mesure μ sur $\mathbb{R}^+$ telle que $\mathbf{f}$ soit la transformée de Stieltjes de μ.

Les conditions techniques.

- f est une fonction analytique sur $\mathbb{C} \setminus \mathbb{R}^+$;
- $\mathrm{Im}(z) > 0$ implique que $\mathrm{Im}(\mathbf{f}(z)) > 0$;
- $\mathrm{Im}(z) > 0$ implique que $\mathrm{Im}(z\mathbf{f}(z)) > 0$;
- $|\mathbf{f}(z)\mathrm{Im}(z)|$ borné sur l' ensemble des complexes z tels que $\mathrm{Im}(z) > 0$,

Propriété

$$\mu(\mathbb{R}^+) = \lim_{y \to \infty} -\mathbf{i}y\mathbf{f}(\mathbf{i}y).$$

La résolvante

On appelle résolvante de la matrice YY^T, la matrice suivante, indexée par $z \in \mathbb{C} \setminus \mathbb{R}^+$:

$$Q(z) = \left(-zI_N + YY^T\right)^{-1}.$$

Il est bien connu en analyse matricielle que la résolvante est un puissant outil pour obtenir des informations sur le spectre d'une matrice donnée.

Propriété La trace normalisée de la résolvante $Q(z)$ de la matrice YY^T est **égale à la transformée de Stieltjes de la mesure spectrale** de YY^T. En effet:

$$\begin{aligned} \mathbf{f}_n(z) &= \frac{1}{N}\mathrm{Trace}(-zI_N + YY^T)^{-1} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} \quad = \quad \int_0^{\infty}\frac{L_N(d\lambda)}{\lambda - z}. \end{aligned}$$

où

$$L_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}.$$

Le modèle i.i.d. centré

Rappelons que dans ce modèle, les entrées de la matrice Y_n sont de la forme:

$$Y_{ij} = \frac{1}{\sqrt{n}} X_{ij},$$

où les X_{ij} sont indépendants et identiquement distribués, centrés, de variance σ^2. Dans ce cas, on a le résultat suivant:

Théorème de Marčenko-Pastur

La mesure spectrale L_N converge vers la probabilité définie par

$$\begin{aligned}\mathbb{P}_{MP}(d\lambda) &= \left(1-\frac{1}{c}\right)\delta_0(d\lambda) \\ &\quad +\frac{1}{c}\frac{\sqrt{(\lambda^+-\lambda)(\lambda-\lambda^-)}}{2\sigma^2\lambda}1_{[\lambda^-,\lambda^+]}(\lambda)d\lambda, \quad (c>1), \\ &= \frac{\sqrt{(\lambda^+-\lambda)(\lambda-\lambda^-)}}{2c\sigma^2\lambda}1_{[\lambda^-,\lambda^+]}(\lambda)d\lambda, \quad (c\leq 1).\end{aligned}$$

où $\lambda^+=(1+\sqrt{c})^2\sigma^2$ et $\lambda^-=(1-\sqrt{c})^2\sigma^2$.

Eléments de preuve

On va caractériser la transformée de Stieltjes limite. Par des arguments de natures différentes,

- manipulations matricielles élémentaires
- perturbation de rang un
- comportement asymptotique de formes quadratiques

on peut montrer que la transdformée de Stieltjes empirique

$$\mathbf{f}_n(z) = \frac{1}{N}\mathrm{Trace}\left(-zI_N + YY^T\right)^{-1}$$

satisfait approximativement l'équation:

$$\mathbf{f}_n(z) \approx \frac{1}{\left(-z + \frac{\sigma^2}{1+\sigma^2\frac{N}{n}\mathbf{f}_n(z)}\right)}.$$

Eléments de preuve -suite (1)-

L'équation-limite suivante émerge naturellement:

$$\mathbf{f}(z) = \frac{1}{\left(-z + \frac{\sigma^2}{1+\sigma^2 c\mathbf{f}(z)}\right)}.$$

C' est un trinôme du second degré:

$$zc\sigma^2\mathbf{f}^2(z) + (z + (c-1)\sigma^2)\mathbf{f}(z) + 1 = 0$$

qui admet deux racines dont **f**:

$$\mathbf{f}(z) \quad = \quad \frac{-(z(c-1)\sigma^2) + \sqrt{(z+(c-1)\sigma^2)^2 - 4c\sigma^2 z}}{2c\sigma^2 z},$$

qui satisfait

$$\lim_{y\to\infty} y|\mathbf{f}(\mathbf{i}y)| = 1$$

et donc est la transformée de Stieltjes d'une mesure de probabilité.

Eléments de preuve -suite (2)-

En utilisant la formule d'inversion d'une transformée de Stieltjes,

$$\mathbb{P}_{MP}([a,b]) = \lim_{y \to 0} \frac{1}{\pi} \int_a^b \mathrm{Im}\mathbf{f}(x + \mathbf{i}y)dx,$$

on peut retrouver la loi de Marčenko-Pastur.

Rappel sur le modèle On s'intéresse au spectre de YY^T lorsque

$$Y = \frac{1}{\sqrt{n}} X + A.$$

On fait de plus l'hypothèse que la mesure spectrale de la matrice déterministe AA^T converge vers une probabilité $\Lambda(d\lambda)$. Alors:

Théorème (Brent Dozier & Silverstein '04) La mesure spectrale de YY^T converge vers une mesure de probabilité dont la transformée de Stieltjes satisfait l'équation implicite suivante:

$$\mathbf{f}(z) = \int \frac{H(\,d\lambda)}{-z(1 + c\sigma^2 \mathbf{f}(z)) + (1-c)\sigma^2 + \frac{\lambda}{1 + c\sigma^2 \mathbf{f}(z)}}.$$

La notion d'équivalent déterministe

Dans le cas d'un modèle à corrélation séparable non centré, i.e.

$$Y = \frac{1}{\sqrt{n}} D^{\frac{1}{2}} X \tilde{D}^{\frac{1}{2}} + A,$$

il **n'existe pas de jeu d'hypothèses simple** garantissant la convergence de la mesure spectrale de YY^T.

Objectif

- construire une quantité déterministe permettant d'approximer la transformée de Stieltjes empirique de YY^T

L'équivalent déterministe pour la TS

Rappelons que la transformée de Stieltjes empirique de YY^* est:

$$\mathbf{f}_n(z) = \frac{1}{N} \mathrm{Trace}\left(-zI_N + YY^*\right)^{-1} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} .$$

Théorème Il existe une **transformée de Stieltjes** déterministe $\mathbf{t}_n(z)$, dépendant des paramètres du modèle D, $\tilde{D}$, A, N et n, telle que

$$\forall z \in \mathbb{C} \setminus \mathbb{R}, \quad \mathbf{f}_n(z) - \mathbf{t}_n(z) \xrightarrow[n \to \infty]{} 0 \quad \text{p.s.}$$

Définition de $\mathbf{t}_n(z)$.

Un système 2×2 Le système suivant admet un unique couple de solutions $(\delta, \tilde{\delta})$

$$\begin{cases} \delta(z) = \dfrac{1}{n}\mathrm{Tr}\left[D\left(-z(I + D\tilde{\delta}) + A(I + \tilde{D}\delta)^{-1}A^T\right)^{-1}\right] \\ \tilde{\delta}(z) = \dfrac{1}{n}\mathrm{Tr}\left[\tilde{D}\left(-z(I + \tilde{D}\delta) + A^T(I + D\tilde{\delta})^{-1}A\right)^{-1}\right] \end{cases}.$$

Définition de $\mathbf{t}_n(z)$.

Un système 2×2 Le système suivant admet un unique couple de solutions $(\delta, \tilde{\delta})$

$$\left\{ \begin{array}{l} \delta(z) = \frac{1}{n}\mathrm{Tr}\left[D \left(-z(I + D\tilde{\delta}) + A(I + \tilde{D}\delta)^{-1}A^T \right)^{-1} \right] \\ \\ \tilde{\delta}(z) = \frac{1}{n}\mathrm{Tr}\left[\tilde{D} \left(-z(I + \tilde{D}\delta) + A^T(I + D\tilde{\delta})^{-1}A \right)^{-1} \right] \end{array} \right. .$$

Un intermédiaire matriciel. On pose

$$T(z) = \left(-z(I + D\tilde{\delta}) + A(I + \tilde{D}\delta)^{-1}A^T \right)^{-1}$$

Définition de $\mathbf{t}_n(z)$.

Un système 2×2 Le système suivant admet un unique couple de solutions $(\delta, \tilde{\delta})$

$$\left\{ \begin{array}{l} \delta(z) = \dfrac{1}{n}\mathrm{Tr}\left[D\left(-z(I + D\tilde{\delta}) + A(I + \tilde{D}\delta)^{-1}A^T\right)^{-1}\right] \\ \\ \tilde{\delta}(z) = \dfrac{1}{n}\mathrm{Tr}\left[\tilde{D}\left(-z(I + \tilde{D}\delta) + A^T(I + D\tilde{\delta})^{-1}A\right)^{-1}\right] \end{array} \right. .$$

Un intermédiaire matriciel. On pose

$$T(z) = \left(-z(I + D\tilde{\delta}) + A(I + \tilde{D}\delta)^{-1}A^T\right)^{-1}$$

Définition de $\mathbf{t}_n(z)$. Alors $\mathbf{t}_n$ est définie par:

$$\mathbf{t}_n(z) = \frac{1}{N}\mathrm{Trace}\, T(z).$$

Expression du **logdet** en fonction de la transformée de Stieltjes

Logdet et transformée de Stieltjes Un calcul élémentaire montre que:

$$I(\rho) = \frac{1}{N}\mathbb{E}\log\det\left(I_N + \frac{Y_n Y_n^T}{\rho}\right). = \mathbb{E}\int_\rho^\infty \left(\frac{1}{\omega} - \mathbf{f}_n(-\omega)\right) d\omega.$$

L'idée .. consiste à remplacer $\mathbf{f}_n$ par son équivalent déterministe $\mathbf{t}_n$.

Théorème Définissions $\bar{I}(\rho)$ par

$$\bar{I}(\rho) = \int_{\rho}^{\infty} \left(\frac{1}{\omega} - \mathbf{t}_n(-\omega) \right) d\omega.$$

Alors

- $I(\rho) - \bar{I}(\rho) \xrightarrow[n \to \infty]{} 0$
- $\bar{I}$ a une expression explicite donnée par:

$$\bar{I}(\rho) = \log\det\left(I + \delta \tilde{D} + \frac{1}{\rho} A^T (I + \tilde{\delta} D)^{-1} A \right) + \log\det\left(I + \tilde{\delta} D \right) - \rho \frac{n}{N} \delta \tilde{\delta}$$

où δ et $\tilde{\delta}$ sont les solutions du sytème 2×2.