Apprentissage statistique et grande dimension

TD 1 : grande dimension

Dans toute la fiche de TD $\|\cdot\|$ désigne la norme usuelle de $\mathbb{R}^p$,

$$\|\boldsymbol{x}\|^2 = \sum_{j=1}^{p} x_j^2, \text{ pour tout } \boldsymbol{x} = (x_1, ..., x_p) \in \mathbb{R}^p$$

et, pour tout $r > 0$, $\boldsymbol{x} \in \mathbb{R}^p$

$$B_p(\boldsymbol{x}, r) := \{\boldsymbol{y} \in \mathbb{R}^p, \|\boldsymbol{x} - \boldsymbol{y}\| \leq r\},$$

la boule de centre x et de rayon r. On notera également $V_p(r)$ le volume de $B_p(r)$. Pour simplifier, $B_p(r) = B_p(0, r)$.

Exercice 1 (Estimateur des moindres carrés et grande dimension)

On suppose que l'on observe des couples $\{(Y_i, \mathbf{x}_i), i = 1, \dots, n\}$ tels que Y_i est la valeur de la variable réponse et $\mathbf{x}_i = (x_i^1, \dots, x_i^d)$ d variables explicatives observées sur le i-ème individu. L'objectif de cet exercice est d'étudier les propriétés de l'estimateur des moindres carrés en grande dimension dans le modèle linéaire gaussien :

$$\mathbf{Y} = \mathbf{X}\beta^* + \boldsymbol{\varepsilon}$$

où l'on rappelle que :

— $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ est un vecteur colonne de taille n,

— $\mathbf{X} = \begin{pmatrix} x_1^1 & \dots & x_1^d \\ \vdots & \vdots & \vdots \\ x_n^1 & \dots & x_n^d \end{pmatrix}$ est une matrice à n lignes et d colonnes,

— $\beta^* = (\beta_1^*, \dots, \beta_d^*)^t$ est un vecteur colonne de taille d, qui est le *paramètre* du modèle et que nous allons chercher à estimer,

— $\boldsymbol{\varepsilon}$ est un vecteur gaussien de moyenne nulle et de matrice de covariance $\sigma^2 I = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2 \end{pmatrix}$.

On supposera dans cet exercice que la matrice $\mathbf{X}$ est non aléatoire (on parle alors de régression linéaire multivariée *à design fixe*).

1. Montrer que toute solution $\overline{\beta}$ du problème d'optimisation
$$\min_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i\beta)^2 \right\}. \tag{1}$$
est solution de l'équation $X^tY = X^tX\overline{\beta}$.
2. La matrice X^tX est-elle toujours inversible ? Que se passe-t'il si elle ne l'est pas ? Dans la suite de l'exercice, nous supposerons que X^tX est inversible et nous noterons $\widehat{\beta}^{(MCO)}$ l'unique solution du problème (**??**).
3. Montrer que $\widehat{\beta}^{(MCO)}$ est un vecteur gaussien dont on précisera la moyenne et la matrice de covariance.
4. Quel est le biais de $\widehat{\beta}^{(MCO)}$?
5. Calculer l'erreur quadratique moyenne de $\widehat{\beta}^{(MCO)}$
$$\mathbb{E}\left[\left\|\widehat{\beta}^{(MCO)} - \beta^*\right\|^2\right].$$
Que se passe-t'il en grande dimension (on pourra regarder le cas $\frac{1}{n}X^tX = I$ par exemple) ?

Réponse de l'exercice 1.

1. Comme la fonction à minimiser $f(\beta) = \frac{1}{n}\sum_{i=1}^n (Y_i - \mathbf{x}_i\beta)^2$ est de classe $\mathcal{C}^1$, toute solution $\overline{\beta}$ du problème d'optimisation $\min_{\beta\in\mathbb{R}^d} f(\beta)$ vérifie l'équation d'Euler $\nabla f(\overline{\beta}) = 0$. Calculons le gradient de f. Nous avons

$$f(\beta) \quad = \quad \frac{1}{n}\sum_{i=1}^n \left(Y_i - \sum_{k=1}^d x_i^k \beta_k\right)^2$$

D'où, pour tout $j = 1, \ldots, d$

$$\begin{aligned}\frac{\partial f}{\partial \beta_j}(\beta) &= -\frac{2}{n}\sum_{i=1}^n x_i^j \left(Y_i - \sum_{k=1}^d x_i^k\beta_k\right) = -\frac{2}{n}\sum_{i=1}^n x_i^j (Y_i - \mathbf{x}_i\beta) = -\frac{2}{n}\left(\left[X^tY\right]_j - \sum_{i=1}^n x_i^j \left[X\beta\right]_i\right) \\ &= -\frac{2}{n}\left(\left[X^tY\right]_j - \left[X^tX\beta\right]_j\right),\end{aligned}$$

où pour un vecteur $\beta = (\beta_1, \ldots, \beta_d)^t \in \mathbb{R}^d$, nous notons $[\beta]_j$ sa j-ème composante β_j et de la même façon, pour $x = (x_1, \ldots, x_n)^t$, nous notons $[x]_i = x_i$ sa i-ème composante. Nous obtenons finalement

$$\nabla f(\beta) = -\frac{2}{n}\left(X^t\mathbf{Y} - X^tX\beta\right).$$

et $\overline{\beta}$ est solution de l'équation d'Euler si et seulement si $X^tY = X^tX\overline{\beta}$.

2. Si X^tX n'est pas inversible, il n'y a plus unicité de la solution : l'équation d'Euler admet soit une infinité de solutions, soit aucune solution. En effet, si une solution $\overline{\beta}$ existe, alors, pour tout $\beta \in \mathrm{Ker}(X^tX) \neq \{0\}$, $\overline{\beta} + \beta$ est aussi solution.
3. Nous avons donc, lorsque X^tX inversible,

$$\widehat{\beta}^{(MCO)} = (X^tX)^{-1}X^tY = (X^tX)^{-1}X^t(X\beta^* + \boldsymbol{\varepsilon}) = \beta^* + (X^tX)^{-1}X^t\boldsymbol{\varepsilon}.$$

Il s'agit donc d'une transformation linéaire du vecteur gaussien $\boldsymbol{\varepsilon}$ (la matrice X n'est pas aléatoire) donc un vecteur gaussien également.

Rappel : si $Z \sim \mathcal{N}(\mu, \Sigma)$ est un vecteur gaussien dans $\mathbb{R}^d$, $b \in \mathbb{R}^r$ et A une matrice de taille $r \times d$, alors $AZ + b \sim \mathcal{N}(A\mu + b, A\Sigma A^t)$.

Comme $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I)$, nous avons

$$\widehat{\beta}^{(MCO)} \sim \mathcal{N}(\beta^*, (X^tX)^{-1}X^t(\sigma^2 I)((X^tX)^{-1}X^t)^t)$$

i.e.

$$\widehat{\beta}^{(MCO)} \sim \mathcal{N}(\beta^*, \sigma^2(X^tX)^{-1}),$$

car $(X^tX)^{-1}X^t(\sigma^2 I)((X^tX)^{-1}X^t)^t = \sigma^2(X^tX)^{-1}X^tX(X^tX)^{-1} = \sigma^2(X^tX)^{-1}$.

4. Par la question précédente $\mathbb{E}[\widehat{\beta}^{(MCO)}] = \beta^*$, donc $\widehat{\beta}^{(MCO)}$ est un estimateur sans biais de β^*.
5. D'après la question précédente

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\beta}^{(MCO)} - \beta^*\right\|^2\right] &= \mathbb{E}\left[\sum_{j=1}^d \left(\widehat{\beta}_j^{(MCO)} - \beta_j^*\right)^2\right] = \sum_{j=1}^d \mathbb{E}\left[\left(\widehat{\beta}_j^{(MCO)} - \mathbb{E}\left[\widehat{\beta}_j^{(MCO)}\right]\right)^2\right] \\ &= \sum_{j=1}^d \mathrm{Var}\left(\widehat{\beta}_j^{(MCO)}\right) = \sum_{j=1}^d \Sigma_{j,j} = \mathrm{tr}(\Sigma),\end{aligned}$$

où $\Sigma = \sigma^2(X^tX)^{-1}$ est la matrice de covariance de $\widehat{\beta}^{(MCO)}$.

Exercice 2 (Loi uniforme en grande dimension)

On considère des vecteurs aléatoires tirés uniformément sur deux ensembles très différents : la boule unité $B_p(1)$ et le pavé $[-1, 1]^p$.

1. Montrer que $V_p(r) = r^p V_p(1)$ (sans calculer $V_p(1)$).
2. Soit $X \sim \mathcal{U}(B_p(1))$ de loi uniforme sur la boule unité, calculer, pour tout $r > 0$, $\mathbb{P}(\|X\| \leq r)$.

3. Soient $X_1, X_2 \sim \mathcal{U}(B_p(1))$ indépendantes.
   (a) Calculer $\mathbb{E}[\|X_1\|^2]$.
   (b) Vérifier que X_i est un vecteur aléatoire centré et en déduire la valeur de $\mathbb{E}[\|X_1 - X_2\|^2]$.
4. Mêmes questions pour $X_1, X_2 \sim \mathcal{U}([-1,1]^p)$.
5. Que pouvez-vous dire sur la loi uniforme en grande dimension pour les deux situations ? Étudier la loi limite et/ou l'espérance des normes des vecteurs.

Réponse de l'exercice 2.

1.
$$V_p(r) = \int_{B_p(r)} dx = \int_{\mathbb{R}^p} 1_{\|x\|\leq r} dx$$

changement de variable $x_j = ry_j$ pour tout $j = 1, \ldots, p$ donc $dx_j = rdy_j$ et $\|x\| = \|ry\| = r\|y\|$.

$$V_p(r) = \int_{\mathbb{R}^p} 1_{r\|y\|\leq r} rdy_1 \ldots rdy_p = r^p \int_{\mathbb{R}^p} 1_{\|y\|\leq 1} dy = r^p V_p(1)$$

2. rappel densité d'une loi uniforme

$$f_X(x) = \frac{1}{V_p(1)} 1_{x\in B_p(1)} = \frac{1}{V_p(1)} 1_{\|x\|\leq 1}$$

Posons $Y = \|X\|$. Déterminons sa fdr

$$\begin{aligned}
P(Y \leq y) &= \int_{\mathbb{R}^p} P(Y \leq y | X = x) f_X(x) dx = \frac{1}{V_p(1)} \int_{\mathbb{R}^p} 1_{\|x\|\leq y} 1_{\|x\|\leq 1} dx \\
&= \frac{\int_{\mathbb{R}^p} 1_{\|x\|\leq y \cap \|x\|\leq 1} dx}{V_p(1)} = \frac{Vol(B_p(1) \cap B_p(y))}{V_p(1)} \\
&= \begin{cases} \frac{Vol(B_p(y))}{V_p(1)} & \text{si } y \leq 1 \\ \frac{Vol(B_p(1))}{V_p(1)} & \text{si } y > 1 \end{cases} = \begin{cases} \frac{y^p V_p(1)}{V_p(1)} & \text{si } y \leq 1 \\ 1 & \text{si } y > 1 \end{cases} = (\min(y,1))^p
\end{aligned}$$

3. (a) Nous avons, par le théorème de Fubini,

$$\mathbb{E}[\|X_1\|^2] = \mathbb{E}\left[\int_0^{+\infty} \mathbf{1}_{\{t\leq\|X_1\|^2\}} dt\right] = \int_0^{+\infty} \mathbb{P}(t < \|X_1\|^2) dt = \int_0^{+\infty} (1 - \mathbb{P}(\sqrt{t} \geq \|X_1\|)) dt$$

En utilisant le résultat de la question 2., cela donne

$$\mathbb{E}[\|X_1\|^2] = \int_0^{+\infty} (1 - \min\{\sqrt{t}, 1\}^p) dt = \int_0^1 (1 - t^{p/2}) dt = 1 - \frac{1}{\frac{p}{2}+1} = \frac{p}{p+2}.$$

   (b) La distribution de X_1 étant symétrique (c'est-à-dire que sa densité est paire), nous avons, par changement de variable $y = -x$,

$$\mathbb{E}[X_1] = \int_{B_p(1)} x f_X(x) dx = \int_{B_p(1)} (-y) f_X(-y) dy = -\int_{B_p(1)} y f_X(y) dy = -\mathbb{E}[X_1].$$

Donc $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. Ensuite

$$\|X_1 - X_2\|^2 = \|X_1\|^2 - 2 < X_1, X_2 > + \|X_2\|^2$$

avec $< x, y >= \sum_i x_i y_i$. Donc

$$E[\|X_1 - X_2\|^2] = E[\|X_1\|^2] - 2E[< X_1, X_2 >] + E[\|X_2\|^2]$$

Or par identique distribution $E[\|X_1\|^2] = E[\|X_2\|^2]$ par indépendance et en supposant que les vecteurs X_i sont centrés.

$$E[< X_1, X_2 >] = \sum_{i=1}^p E[X_{1,i} X_{2,i}] = \sum_{i=1}^p E[X_{1,i}] E[X_{2,i}] = 0$$

En utilisant le résultat de la question a),

$$\mathbb{E}[\|X_1 - X_2\|^2] = 2\mathbb{E}[\|X_1\|^2] = \frac{2p}{p+2}.$$

4. (a) Pour utiliser $E[\|X_1 - X_2\|^2] = 2E[\|X_1\|^2]$, vérifions que $X_1, X_2 \overset{iid}{\sim} X$ avec $X \sim \mathcal{U}([-1,1]^p)$ sont bien des variables centrées.
La densité du vecteur X est plus simple

$$f_X(x) = \frac{1}{Vol([-1,1]^p)} 1_{[-1,1]^p}(x) = \frac{1}{2^p} 1_{[-1,1]^p}(x) = \frac{1}{2} 1_{[-1,1]}(x_1) \times \cdots \times \frac{1}{2} 1_{[-1,1]}(x_p)$$

Donc toutes les composantes sont indépendantes contrairement au cas précédent : $f_{X_i}(x_i) = \frac{1}{2} 1_{[-1,1]}(x_i)$.
La fdr s'obtient facilement par intégration

$$F_X(x) = \prod_{i=1}^{p} \frac{\min(x_i + 1, 2)}{2} 1_{[-1,+\infty[^p}(x)$$

Le calcul d'espérance est nettement plus simple $E[\|X_1 - X_2\|] = 2E[\|X_1\|^2]$. car les vecteurs X_1, X_2 sont iid et centrés.

$$E[\|X_1\|^2] = \sum_{i=1}^{p} E[X_{1,i}^2] = \sum_{i=1}^{p} \int_{-1}^{1} x_i^2 \frac{1}{2} 1_{[-1,1]}(x_i) dx_i = \sum_{i=1}^{p} \left[x^3/6\right]_{-1}^{1} = \sum_{i=1}^{p} 1/3 = p/3$$

Donc $E[\|X_1 - X_2\|^2] = 2p/3$.

(b) Le calcul d'espérance est nettement plus simple

$$E(X_i) = \int_{-1}^{1} x_i \frac{1}{2} 1_{[-1,1]}(x_i) dx_i = 0 \Rightarrow E[X] = 0_{\mathbb{R}^p}$$

5. (a) Analysons
— la fdr

$$\begin{cases} \forall y < 1, P(\|X\| \leq y) = y^p \underset{p\to+\infty}{\longrightarrow} 0, \\ \forall y \geq 1, P(\|X\| \leq y) = 1 \underset{p\to+\infty}{\longrightarrow} 1, \end{cases} \Leftrightarrow \|X\| \underset{p\to+\infty}{\xrightarrow{\mathcal{L}}} 1_{\|X\|=1}$$

C'est à dire les points se retrouvent sur la sphère unité
— la norme 2 moyenne

$$E[\|X_1\|^2] = \frac{p}{p+2} \underset{p\to+\infty}{\longrightarrow} 1, E[\|X_1 - X_2\|^2] = 2\frac{p}{p+2} \underset{p\to+\infty}{\longrightarrow} 2.$$

La distance entre deux points aléatoire tend vers la distance maximale entre deux points de la sphère unité.

(b) Analysons
— la fdr (trop complexe)
— la norme 2 en moyenne

$$E[\|X_1\|]^2 = 2p/3 \underset{p\to+\infty}{\longrightarrow} +\infty, E[\|X_1 - X_2\|]^2 = 4p/3 \underset{p\to+\infty}{\longrightarrow} +\infty,$$

La distance d'un point à l'origine tends vers l'infini.

Exercice 3 (Volume de la boule unité en dimension p)

On étudie en détail le volume de la boule unité en grande dimension.

1. Montrer que $V_1(1) = 2$ et $V_2(1) = \pi$.
2. Prouver que, pour $p \geq 3$, $V_p(1) = \frac{2\pi}{p} V_{p-2}(1)$.
3. En déduire que, pour tout $k \in \mathbb{N}\backslash\{0\}$,

$$V_{2k}(1) = \frac{\pi^k}{k!}$$

et, pour tout $k \in \mathbb{N}$,

$$V_{2k+1}(1) = \frac{2^{2k+1} k! \pi^k}{(2k+1)!}.$$

4. Donner un équivalent de $V_p(r)$ quand $p \to \infty$. On rappelle la formule de Stirling

$$n! \sim_{n\to\infty} \sqrt{2\pi n} n^n e^{-n}.$$

5. Nous souhaitons couvrir l'hypercube $[0,1]^p$ par des boules de rayon 1 c'est-à-dire trouver $x_1, ..., x_n \in \mathbb{R}^p$ tels que

$$[0,1]^p \subset \bigcup_{i=1}^{n} B_p(x_i, 1). \tag{2}$$

Notons n_p le nombre minimal de points nécéssaire pour couvrir $[0,1]^p$.

(a) Montrer que $n_p = 1$ lorsque $p \leq 4$. Que se passe-t'il pour $p = 5$?

(b) Montrer que si (**??**) est vérifiée alors $n_p V_p(1) \geq 1$.

Réponse de l'exercice 3.

1.

$$V_1(1) = \int_{\mathbb{R}} 1_{\|x_1\|\leq 1} dx_1 = \int_{-1}^{1} dx_1 = 2$$

changement en coordonnées polaires

$$V_2(1) = \int_{\mathbb{R}^2} 1_{\|x\|\leq 1} dx_1 dx_2 = \int_0^\infty \int_{-\pi}^{\pi} 1_{\rho \leq 1} \rho d\rho d\theta = \int_0^1 \rho d\rho 2\pi = \pi$$

2. pour $p \geq 3$

$$\begin{aligned} V_p(1) &= \int_{\mathbb{R}^p} 1_{x_1^2+\cdots+x_p^2\leq 1} dx_1 dx_2 dx_3 \ldots dx_p \\ &= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^{p-2}} 1_{\sqrt{x_3^2+\cdots+x_p^2}\leq \sqrt{1-x_1^2-x_2^2}} dx_3 \ldots dx_p \right) 1_{x_1^2+x_2^2\leq 1} dx_1 dx_2 \\ &= \int_{\mathbb{R}^2} V_{p-2}(\sqrt{1-x_1^2-x_2^2}) 1_{x_1^2+x_2^2\leq 1} dx_1 dx_2 \end{aligned}$$

Comme $V_p(r) = r^p V_p(1)$ et par changement en coordonnées polaires

$$\begin{aligned} V_p(1) &= V_{p-2}(1) \int_{\mathbb{R}^2} (1-x_1^2-x_2^2)^{p/2-1} 1_{x_1^2+x_2^2\leq 1} dx_1 dx_2 \\ &= V_{p-2}(1) \int_0^1 (1-\rho^2)^{p/2-1} \rho d\rho \int_{-\pi}^{\pi} d\theta = V_{p-2}(1) 2\pi \int_0^1 (-2)\rho(1-\rho^2)^{p/2-1} d\rho/(-2) \\ &= -V_{p-2}(1)\pi \left[\frac{(1-\rho^2)^{p/2}}{p/2} \right]_0^1 = \frac{2\pi}{p} V_{p-2}(1) \end{aligned}$$

3. Pour p pair on a donc

$$V_p(1) = \frac{2\pi}{p} V_{p-2}(1) = \frac{2\pi}{p}\frac{2\pi}{p-2} V_{p-4}(1) = \frac{2\pi}{p}\frac{2\pi}{p-2} \cdots \frac{2\pi}{4} V_2(1) = \frac{\pi^{p/2-1}}{p/2(p/2-1)\ldots 2}\pi = \frac{\pi^{p/2}}{(p/2)!}$$

Pour p impair on a donc

$$V_p(1) = \frac{2\pi}{p} V_{p-2}(1) = \frac{2\pi}{p}\frac{2\pi}{p-2} \cdots \frac{2\pi}{3} V_1(1) = \frac{2^{(p+1)/2}\pi^{(p-1)/2}}{p(p-2)\ldots 3} = \frac{2^{(p+1)/2}\pi^{(p-1)/2}}{p!}(p-1)\ldots 4 \times 2$$

Or $(p-1)\ldots 4 \times 2 = 2^{(p-1)/2} \frac{(p-1)}{2} \ldots 2 \times 1$

$$V_p(1) = \frac{2^p \pi^{(p-1)/2}}{p!} \left(\frac{(p-1)}{2} \right)!$$

4. formule de Stirling k pair

$$V_{2k}(r) \sim r^{2k} \frac{\pi^k}{\sqrt{2\pi k} k^k} e^k = \frac{r^{2k}}{\sqrt{2}} \left(\frac{\pi}{k}\right)^{k-1/2} e^k = \frac{1}{\sqrt{2}} \left(e\frac{r^2\pi}{k} \right)^k \left(\frac{\pi}{k}\right)^{-1/2} \underset{k\to+\infty}{\longrightarrow} 0$$

pour k impair

$$\begin{aligned}
V_{2k+1}(1) &\sim r^{2k+1}\frac{2^{2k+1}\pi^k\sqrt{2\pi k}k^ke^{-k}}{\sqrt{2\pi(2k+1)}(2k+1)^{2k+1}}e^{2k+1} = r^{2k+1}\left(\frac{k}{2k+1}\right)^{k+1/2}\frac{2^{2k+1}\pi^k}{(2k+1)^{k+1}}e^{k+1}\\
&\sim r^{2k+1}\left(\frac{1}{2}\right)^{k+1/2}\frac{4^{k+1/2}\pi^k}{(2k+1)^{k+1}}e^{k+1} = r^{2k+1}\frac{2^{k+1/2}(e\pi)^k}{(2k+1)^{k+1}}e^1\\
&\sim \left(\frac{r^2e\pi}{k}\right)^k\frac{e^1\sqrt{2}}{k}\underset{k\to+\infty}{\longrightarrow} 0
\end{aligned}$$

5. (a) Considérons $p=1$

$$[0,1]\subset B_1(0,1)=[-1,1]$$

Considérons $p=2$

$$[0,1]^2\subset B_2(\begin{pmatrix}1/2\\1/2\end{pmatrix},1)$$

Considérons p quelconque et la boule de centre $x=1/2$, c'est à dire

$$B_p\left(x,1\right)=B_p\left(\begin{pmatrix}1/2\\ \vdots\\ 1/2\end{pmatrix},1\right)$$

$$\forall y\in[0,1]^p,(1/2-y_j)^2\le(1/2)^2\Rightarrow \|x-y\|_p^2=\sum_{j=1}^p(1/2-y_j)^2\le\sum_{j=1}^p(1/2)^2=p/4$$

Donc pour $p\le 4$ on a pour tout $y\in[0,1]^p$, $\|x-y\|_p^2\le 1$. Pour tout $p\ge 5$, $\|x\|=5/4>1$ donc $0\notin B_p\left(\begin{pmatrix}1/2\\ \vdots\\ 1/2\end{pmatrix},1\right)$.

(b) Si $[0,1]^p\subset\bigcup\limits_{i=1}^n B_p(x_i,1)$ alors

$$Vol\left([0,1]^p\right)\le Vol\left(\bigcup_{i=1}^n B_p(x_i,1)\right)\Leftrightarrow 1\le\sum_{i=1}^p V_p(1)=n_pV_p(1)$$

Autrement dit

$$n_p\ge\frac{1}{V_p(1)}\underset{p\to+\infty}{\longrightarrow}+\infty$$

Exercice 4 (Loi normale en grande dimension)

On s'intéresse à la loi normale centrée réduite en dimension p dont on rappelle la densité.

$$g_p(x)=\frac{1}{\sqrt{2\pi}^p}e^{-\|x\|^2/2}.$$

1. Quel est le maximum de g_p (le mode de la distribution) ? que se passe-t-il quand $p\to+\infty$?
2. Nous souhaitons déterminer la masse de la "cloche" de la gaussienne c'est-à-dire la partie de plus grande densité. Soit $\delta>0$ proche de 0, nous notons

$$B_{p,\delta}:=\left\{x\in\mathbb{R}^p,g_p(x)\ge\delta g_p(0)\right\}.$$

   (a) Faire un dessin en dimension $p=1$.
   (b) Faire un dessin en dimension $p=2$.
   (c) À l'aide de l'inégalité de Markov, montrer que

$$\mathbb{P}\left(X\in B_{p,\delta}\right)\le\delta^{-1}2^{-p/2}.$$

   (d) Conclure sur l'effet de la dimensions sur la loi normale centrée réduite.

1.

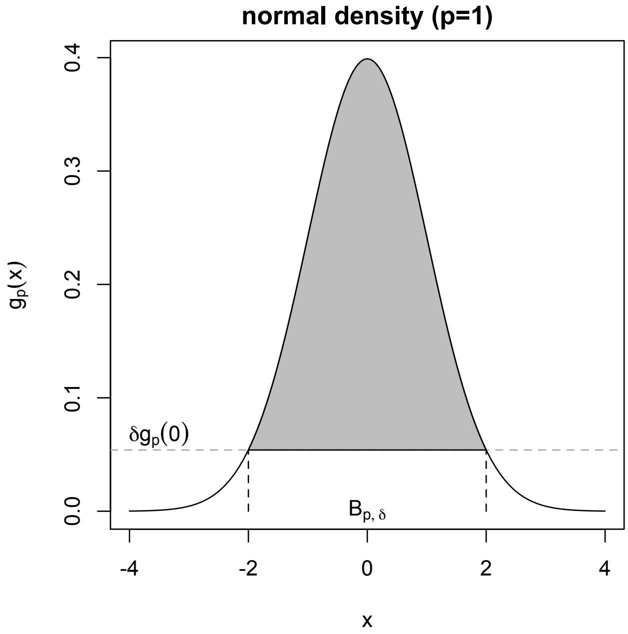
$$\max_x g_p(x) = \max_x \exp(-1/2 \sum_{i=1}^{p} x_i^2) = \min_x 1/2 \sum_{i=1}^{p} x_i^2$$

Le mode de la distribution est atteint en $x = 0$. Le mode a pour valeur

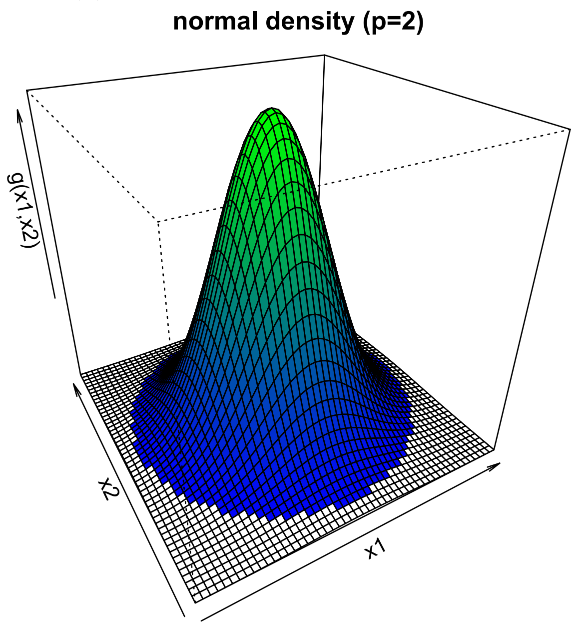
$$g_p(0) = (2\pi)^{-p/2} \underset{p \to +\infty}{\longrightarrow} 0$$

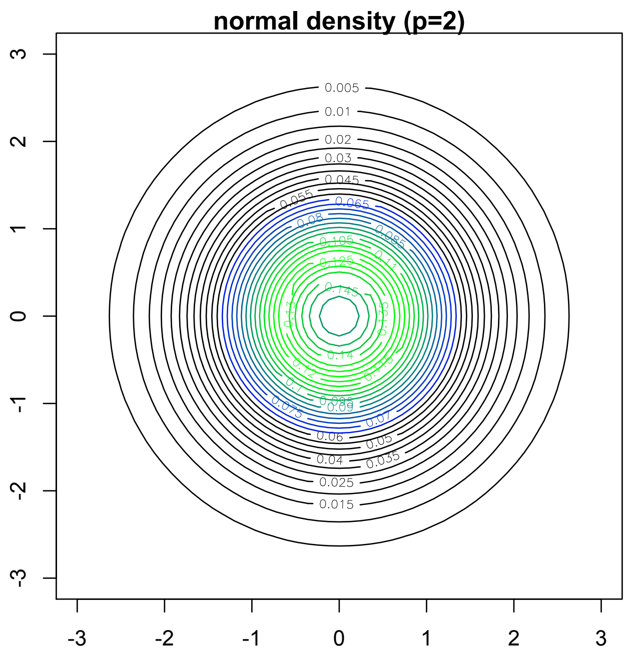
Donc la distribution devient extremement plate en grande dimension.

2. (a) En dimension 1



(b) En dimension 2





(c) Rappel de l'inégalité : $\forall a > 0, P(Z \geq a) \leq E[Z]/a$ pour Z une variable positive p.s..

$$\begin{aligned}
P(X \in B_{p,\delta}) &= P(g_p(X) \geq \delta g_p(0)) = P(\frac{1}{\sqrt{2\pi}^p} e^{-\|X\|^2/2} \geq \delta(2\pi)^{-p/2}) \\
&= P(\exp(-\|X\|^2/2) \geq \delta) \\
&\leq E(\exp(-\|X\|^2/2))/\delta = 1/\delta \int_{\mathbb{R}^p} \exp(-1/2 \sum_{i=1}^{p} x_i^2) \frac{1}{\sqrt{2\pi}^p} e^{-\|x\|^2/2} dx \\
&\leq 1/\delta \int_{\mathbb{R}^p} \frac{1}{\sqrt{2\pi}^p} e^{-\|x\|^2} dx = \frac{1}{\delta\sqrt{2^p}} \int_{\mathbb{R}^p} \frac{1}{\sqrt{2\pi}^p} e^{-\|y\|^2/2} dy \\
&\leq \frac{1}{\delta\sqrt{2^p}} \underset{p \to +\infty}{\longrightarrow} 0
\end{aligned}$$

La probabilité d'être dans cette zone $B_{p,\delta}$ tend exponentiellement vite vers 0. Donc la probabilité d'être dans la queue de distribution tend exponentiellement vite vers 1.