

Vecteurs Gaussiens et théorèmes de convergence

Prépa Agreg Option A

Nathan Huguenin

1 Introduction

Largement inspiré du livre de M-L. Chabanol et J-J. Ruch.

2 Vecteurs Gaussiens

2.1 Vecteurs aléatoires

Definition 2.1. Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ un vecteur aléatoire de $\mathbb{R}^d$ tel que $\mathbb{E}[X_i^2] < \infty$ pour tout i . L'espérance de X est le vecteur

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix}.$$

Sa matrice de covariance est

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] = \begin{pmatrix} \text{Var}[X_1] & \dots & \text{Cov}[X_1, X_d] \\ \vdots & \ddots & \vdots \\ \text{Cov}[X_d, X_1] & \dots & \text{Var}[X_d] \end{pmatrix}$$

La covariance de X est finie grâce à l'hypothèse de carré intégrable et l'inégalité de Cauchy-Schwarz. Si X est un vecteur aléatoire de $\mathbb{R}^d$ de moyenne m et de matrice de covariance M , et si A est une matrice réelle de taille $d' \times d$, alors le vecteur aléatoire AX de $\mathbb{R}^{d'}$ a pour moyenne Am et pour matrice de covariance $AMAT^T$ (il suffit de calculer). Le théorème/définition suivant caractérise les matrices de covariance et fournit des propriétés importantes de celles-ci.

Definition 2.2. Une matrice symétrique $M \in M_d(\mathbb{R})$ est dite semi-définie positive ssi elle vérifie l'une de ces propriétés équivalentes :

1. $V^T M V \geq 0$ pour tout vecteur $V \in \mathbb{R}^d$ (i.e. la forme bilinéaire associée à M est positive)
2. $\text{Sp}(M) \subset [0, \infty[$ (i.e. les valeurs propres de M sont positives ou nulles)
3. Il existe une matrice réelle A telle que $AA^T = M$ (pas unique en général)
4. M est la matrice de covariance d'un vecteur aléatoire de $\mathbb{R}^d$.

Proof. • $1 \Rightarrow 2$: on applique 1 à un vecteur propre x pour la valeur propre λ : on a $0 \leq x^T M x = \lambda |x|^2$, donc $\lambda \geq 0$.

- $2 \Rightarrow 3$: par le théorème spectral il existe une matrice orthogonale P telle que $M = P D P^T$ avec $D = \text{Diag}(\lambda \in \text{Sp}(M))$. Comme D est positive, il suffit de prendre $A = P \text{Diag}(\sqrt{\lambda}, \lambda \in \text{Sp}(M))$.

- 3 $\Rightarrow$ 4 : soit X un vecteur aléatoire de $\mathbb{R}^d$ dont les composantes sont indépendantes, centrées et de variance 1. On a donc $\mathbb{E}[X] = 0$ et $\text{Var}[X] = I_d$. Le vecteur AX est alors centré, de matrice de covariance $AI_dA^T = AA^T = M$.
- 4 $\Rightarrow$ 1 : une matrice de covariance est symétrique. De plus si M est la matrice de covariance d'un vecteur aléatoire X de $\mathbb{R}^d$ et si V est un vecteur de $\mathbb{R}^d$, $V^T M V = \text{Var}(v_1 X_1 + \dots + v_d X_d) \geq 0$. $\square$

On rappelle également la définition de la fonction caractéristique.

Definition 2.3. La fonction caractéristique d'un vecteur aléatoire X de $\mathbb{R}^d$ est la fonction $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par

$$\Phi_X(u) := \mathbb{E}[e^{iu^T X}] = \mathbb{E}[e^{i\langle u, X \rangle}]$$

pour $u \in \mathbb{R}^d$.

La fonction caractéristique caractérise la loi.

2.2 Vecteurs Gaussiens

2.2.1 Généralités

Definition 2.4. Un vecteur aléatoire de $\mathbb{R}^d$ est dit *Gaussien* si toute combinaison linéaire de ses composantes est une variable Gaussienne réelle.

Exercice 2.5. Montrer qu'un vecteur aléatoire dont toutes les composantes sont indépendantes et Gaussiennes est un vecteur Gaussien. En revanche, il ne suffit pas que les composantes soient Gaussiennes pour construire un vecteur Gaussien : soit Y de loi $\mathcal{N}(0, 1)$ et ε indépendante de loi de Rademacher $\frac{1}{2}(\delta_1 + \delta_{-1})$. Montrer que Y et εY sont Gaussiennes, mais que le vecteur $(Y, \varepsilon Y)$ n'est pas Gaussien. Montrer de plus que Y et εY ne sont pas indépendantes alors que leur covariance est nulle.

Proof. On peut remarquer que $\mathbb{P}(Y + \varepsilon Y = 0) = \frac{1}{2}$ donc la variable $Y + \varepsilon Y$ n'est pas Gaussienne. $\square$

Proposition 2.6. Un vecteur aléatoire X de $\mathbb{R}^d$ est un vecteur Gaussien de moyenne m et de matrice de covariance Γ si et seulement si sa fonction caractéristique est donnée par

$$\Phi_X(u) = \exp\left(iu^T m - \frac{1}{2}u^T \Gamma u\right).$$

La loi d'un vecteur Gaussien est donc caractérisée par sa moyenne et sa matrice de covariance, on la note $\mathcal{N}(m, \Gamma)$.

Proof. C'est la définition 2.4 et la définition de la fonction caractéristique d'une Gaussienne réelle. $\square$

Proposition 2.7. Soit Γ une matrice de covariance de taille $d \times d$, m un vecteur de $\mathbb{R}^d$ et Z un vecteur Gaussien standard (i.e. de loi $\mathcal{N}(0, I_d)$). Soit A une racine carrée matricielle de Γ (i.e. $AA^T = \Gamma$). Alors le vecteur $AZ + m$ suit la loi $\mathcal{N}(m, \Gamma)$. En particulier, la loi $\mathcal{N}(0, I_d)$ est invariante par transformation orthogonale.

Proof. Comparer les fonctions caractéristiques. $\square$

Proposition 2.8. Toute matrice symétrique semi-définie positive est la matrice de covariance d'un vecteur Gaussien.

Proof. Dans la preuve du point 3 $\Rightarrow$ 4 de la définition 2.2 il suffit de prendre X dont toutes les composantes sont des Gaussiennes centrées réduites indépendantes. $\square$

Exercice 2.9 (Vecteurs Gaussiens à densité). *Montrer le résultat suivant : la loi $\mathcal{N}(m, \Gamma)$ sur $\mathbb{R}^d$ admet une densité par rapport à la mesure de Lebesgue si et seulement si Γ est inversible, et dans ce cas sa densité est donnée par*

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det \Gamma}} \exp \left(-\frac{1}{2}(x - m)^T \Gamma^{-1}(x - m) \right).$$

Proof. La loi $\mathcal{N}(m, \Gamma)$ est la loi du v.a. $AZ + m$ où Z a toutes ses composantes Gaussiennes centrées réduites indépendantes et où A est une racine carrée de Γ . Z a pour densité

$$f_Z(z) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}z_i^2 \right) = \frac{1}{\sqrt{(2\pi)^d}} \exp \left(-\frac{1}{2}|z|^2 \right).$$

Soit h une fonction continue bornée. On a

$$\mathbb{E}[h(X)] = \mathbb{E}[h(AZ + m)] = \int_{\mathbb{R}^d} h(Az + m) f_Z(z) dz.$$

Le changement de variable $x = Az + m$ est un difféomorphisme si et seulement si A est inversible, c'est-à-dire si et seulement si Γ est inversible (et dans ce cas $\Gamma^{-1} = (A^{-1})^T A^{-1}$). Le Jacobien est donné par $\det(A^{-1}) = 1/\sqrt{\det \Gamma}$. Ainsi

$$\mathbb{E}[h(X)] = \frac{1}{\sqrt{(2\pi)^d \det \Gamma}} \int_{\mathbb{R}^d} h(x) \exp \left(-\frac{1}{2}(x - m)^T \Gamma^{-1}(x - m) \right) dx.$$

□

Proposition 2.10. *Pour tout vecteur Gaussien X de $\mathbb{R}^d$, les assertions suivantes sont équivalentes :*

1. $X_1, \dots, X_d$ sont mutuellement indépendantes.
2. $X_1, \dots, X_d$ sont deux à deux indépendantes.
3. La matrice de covariance de X est diagonale.

Proof. Il suffit de montrer $3 \Rightarrow 1$ (les autres implications sont évidentes). On note la matrice de covariance de X $\Gamma = \text{Diag}(\Gamma_{11}, \dots, \Gamma_{dd})$ et m sa moyenne. Alors pour tout $u \in \mathbb{R}^d$,

$$\Phi_X(u) = \exp \left(iu^T m - \frac{1}{2}u^T \Gamma u \right) = \exp \left(i \sum_{k=1}^d u_k m_k - \frac{1}{2} \sum_{k=1}^d \Gamma_{kk} u_k^2 \right) = \prod_{k=1}^d \exp \left(iu_k m_k - \frac{1}{2} \Gamma_{kk} u_k^2 \right) = \prod_{k=1}^d \Phi_{X_k}(u_k).$$

□

2.2.2 Théorème de Cochran

On continue cette section par le théorème de Cochran, qui est une sorte de théorème de Pythagore pour les vecteurs Gaussiens, et qui est notamment à la base des méthodes de régression linéaire qui seront étudiées plus tard. La loi du χ^2 à d degrés de libertés est la loi de la norme au carré d'un vecteur Gaussien standard de $\mathbb{R}^d$. Plus précisément,

Definition 2.11. *Soient $X_1, \dots, X_d$ des variables i.i.d. de loi $\mathcal{N}(0, 1)$. La loi du χ^2 à d degrés de libertés, notée $\chi^2(d)$, est la loi de la variable*

$$\sum_{k=1}^d X_k^2.$$

Sa densité est définie par

$$f_{\chi^2(d)}(x) = \frac{1}{2\Gamma(d/2)} \left(\frac{x}{2} \right)^{d/2-1} \exp \left(-\frac{x}{2} \right) \mathbb{1}_{x>0}.$$

et sa fonction caractéristique est donnée par

$$\Phi_{\chi^2(d)}(u) = \frac{1}{(1 - 2iu)^{d/2}}.$$

Pour rappel, la fonction Gamma d'Euler est définie sur $\{Re(z) > 0\}$ par $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ puis sur $\mathbb{C}$ par prolongement analytique. Elle n'a pas de zéros, et est méromorphe sur $\mathbb{C}$ avec des pôles simples aux entiers négatifs. Elle vérifie $\Gamma(n) = (n-1)!$ pour $n \geq 1$. On peut maintenant énoncer le théorème de Cochran.

Theorem 2.12 (Théorème de Cochran). *Soit X un vecteur Gaussien de $\mathbb{R}^d$ de loi $\mathcal{N}(m, \sigma^2 I_d)$ avec $\sigma^2 > 0$. Soit $\mathbb{R}^d = E_1 \oplus \dots \oplus E_p$ une décomposition en somme directe de $\mathbb{R}^d$ en p s.e.v. orthogonaux de dimensions $d_1, \dots, d_p$ (donc $d_1 + \dots + d_p = d$). Pour $k = 1, \dots, p$, on note Π_k la projection orthogonale sur E_k et $Y_k = \Pi_k X$. Alors :*

1. $Y_1, \dots, Y_p$ sont des vecteurs Gaussiens indépendants et pour $k = 1, \dots, p$, Y_k est de loi $\mathcal{N}(\Pi_k m, \sigma^2 \Pi_k)$.
2. Les variables $|Y_1 - \Pi_1 m|^2, \dots, |Y_p - \Pi_p m|^2$ sont indépendantes et pour $k = 1, \dots, p$, $\frac{1}{\sigma^2} |Y_k - \Pi_k m|^2$ est de loi $\chi^2(d_k)$.

Proof. La matrice de covariance de X est diagonale donc ses composantes sont indépendantes, donc les Y_k sont des vecteurs Gaussiens indépendants. La loi de Y_k est la loi $\mathcal{N}(\Pi_k m, \Gamma)$ avec $\Gamma = \Pi_k \sigma^2 I_d \Pi_k^T = \sigma^2 \Pi_k \Pi_k = \sigma^2 \Pi_k$. Ainsi, les variables $|Y_k - \Pi_k m|^2$ sont indépendantes, et

$$\frac{1}{\sigma^2} |Y_k - \Pi_k m|^2 = \sum_{i=1}^{d_k} \left(\frac{Y_{k,i} - (\Pi_k m)_i}{\sigma} \right)^2$$

qui est la somme de d_k carrés de v.a. normales centrées réduites indépendantes, donc de loi $\chi^2(d_k)$. $\square$

Voici un premier corollaire du théorème de Cochran, utile en estimation statistique, qui sera revu plus tard. D'autres applications seront aussi abordées dans le chapitre sur la régression linéaire.

Corollary 2.13. *Soit $(X_1, \dots, X_n)$ des variables i.i.d. de loi $\mathcal{N}(m, \sigma^2)$. Soient*

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \quad ; \quad \hat{S}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Alors les variables $\bar{X}_n$ et $\hat{S}_n^2$ sont indépendantes, de lois respectives $\mathcal{N}(m, \sigma^2/n)$ et $\chi^2(n-1)$.

Exercice 2.14 (Preuve du corollaire). *En appliquant le théorème de Cochran au sous-espace vectoriel $F = \text{Vect}(\mathbb{1}_n)$ où $\mathbb{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$ et au vecteur Gaussien $Y = \frac{1}{\sigma}(X - m\mathbb{1}_n)$, prouver le corollaire.*

Proof. Y ainsi défini suit la loi $\mathcal{N}(0, I_n)$.

- La matrice de la projection orthogonale sur F est $\Pi_F = \mathbb{1}_n (\mathbb{1}_n^T \mathbb{1}_n)^{-1} \mathbb{1}_n^T = \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T$.
- La matrice de la projection orthogonale sur $F^\perp$ est $\Pi_{F^\perp} = I_n - \Pi_F$.
- $\Pi_F Y = \frac{1}{\sigma} (\Pi_F X - m \Pi_F \mathbb{1}_n) = \frac{1}{\sigma} (\bar{X}_n - m) \mathbb{1}_n$.
- $\Pi_{F^\perp} Y = Y - \Pi_F Y = \frac{1}{\sigma} (X - \bar{X}_n \mathbb{1}_n)$, donc $|Y - \Pi_F Y|^2 = \frac{n-1}{\sigma^2} \hat{S}_n^2$. On conclut grâce au théorème de Cochran.

$\square$

2.2.3 Simulation des vecteurs Gaussiens

On a déjà vu les méthodes de type Box-Muller pour générer des Gaussiennes indépendantes. Les propriétés des vecteurs Gaussiens qu'on a énoncé suggèrent une méthode pour générer des vecteurs Gaussiens de moyenne et de matrice de covariance données, à condition de savoir calculer une racine carrée matricielle de la matrice de covariance.

Exercice 2.15. 1. Essayer la fonction `rd.normal`, est-elle suffisante pour générer des vecteurs Gaussiens quelconques ?

2. Soit $X = (X_1, \dots, X_n)^T$ un vecteur formé de Gaussiennes standard i.i.d. Soit A une matrice réelle carrée de taille n . Quelle est la loi de AX ?

3. Comment construire un vecteur Gaussien de loi $\mathcal{N}(0, \Gamma)$ où Γ est réelle symétrique semi-définie positive ?

4. En utilisant la fonction `np.linalg.cholesky`, écrire une routine permettant de simuler un nombre donné de vecteurs Gaussiens de moyenne et de matrice de covariance données.

5. En simulant un grand nombre de vecteurs Gaussiens de petite taille de moyenne et de covariance connues, tester la justesse de la routine construite.

Proof. La fonction `rd.normal` permet de simuler des vecteurs Gaussiens dont les composantes sont i.i.d. Si A est une racine carrée matricielle d'une matrice de covariance Γ , alors on peut passer d'un vecteur Gaussien standard à un vecteur Gaussien de covariance Γ en multipliant par A . Dans le cas où la matrice de covariance est définie positive, on peut utiliser la décomposition de Cholesky pour calculer A . Le code suivant produit un vecteur Gaussien de moyenne et de covariance donnée dans ce cas-là.

```
def vectGauss(m,gamma,dim):
    x = rd.normal(0,1,dim)
    A = np.linalg.cholesky(gamma)
    xt = x.reshape(-1,1)
    return A@xt + m
```

La valeur -1 signifie qu'on laisse Python adapter la dimension du nouveau vecteur en fonction du nombre d'entrées de l'ancien. On peut aussi utiliser `xt = x[... ,None]`. Le `@` est un raccourci pour le produit matriciel `np.dot`. Si la matrice de covariance n'est pas définie positive, alors la décomposition de Cholesky existe mais n'est pas gérée par Python. On peut alors décomposer la matrice dans une base orthonormée et définir la racine carrée comme dans la démo du poly. Le code suivant montre comment coder ceci.

```
def vectGauss(m,gamma,dim):
    x = rd.normal(0,1,dim)
    xt = x.reshape(-1,1)
    evals, evecs = np.linalg.eigh(gamma)
    evals = np.diag(np.sqrt(evals))
    return (evecs@evals@xt + m
```

□

2.2.4 Autres exercices

Exercice 2.16. Soit (X, Y) de loi $\mathcal{N}(0, I_2)$. Soit $U = \frac{1}{\sqrt{2}}(X + Y)$ et $V = \frac{1}{\sqrt{2}}(X - Y)$. Montrer que (U, V) est un vecteur Gaussien et calculer sa covariance.

Proof. C'est un fait général : (U, V) est l'image de (X, Y) par une transformation orthogonale, donc c'est un vecteur Gaussien standard (si A est une matrice orthogonale, alors $AI_dA^T = I_d$). □

Exercice 2.17. Soit X de loi $\mathcal{N}(m, \Gamma)$. Déterminer la loi de $(X - m)^T \Gamma^{-1} (X - m)$.

Proof. Ecrire $Z = \Gamma^{-\frac{1}{2}}(X - m)$, Z est standard donc la loi recherchée est la loi du χ^2 . □

Exercice 2.18. Soit X de loi $\mathcal{N}(m, \Gamma)$ dans $\mathbb{R}^3$ avec

$$m = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad ; \quad \Gamma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix}.$$

1. Déterminer $\ker \Gamma$.

2. Soit

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Donner les caractéristiques du vecteur Gaussien AX . En déduire que $X_2 - X_3 = 1$ p.s.

3. Le vecteur (X_1, X_2) admet-il une densité dans $\mathbb{R}^2$? Si oui, laquelle ?

4. Donner le support de la loi de X .

Proof. 1. En résolvant un système on trouve $\ker \Gamma = \text{Vect}(0, 1, -1)$.

2. On a $\mathbb{E}[AX] = Am = m$, $\text{Var}[AX] = A\Gamma A^T = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}$. De plus $AX = (X_1, X_2 - X_3, X_3)$, donc $\text{Var}[X_2 - X_3] = 0$, donc $X_2 - X_3 = \mathbb{E}[X_2 - X_3] = 1$ p.s.

3. La matrice de covariance de (X_1, X_2) est la matrice extraite de $\Gamma \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, qui est inversible de déterminant 2. Donc (X_1, X_2) admet une densité donnée par la formule du cours.

4. De manière générale, soit Γ la matrice de covariance de X de taille $n \times n$ non inversible et soit K son noyau de dimension $< n$. Alors pour tout vecteur $v \in K$, $\Gamma v = 0$. Ainsi,

$$\text{Var}\left(\sum_i a_i X_i\right) = \sum_{i,j} a_j \text{Cov}(X_i, X_j) a_i = a^T \Gamma a = 0$$

Ainsi, il existe des combinaisons linéaires des composantes de X constantes p.s. Plus précisément, soit $\{a_1, \dots, a_k\}$ une base de K avec $k < n$, alors il existe $(\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$ t.q. $\mathbb{P}(a_1^T X = \alpha_1, \dots, a_k^T X = \alpha_k) = 1$. Le support de la loi de X est donc le sous-espace affine $H = \{x \in \mathbb{R}^n \text{ t.q. } a_1^T x = \alpha_1, \dots, a_k^T x = \alpha_k\}$. □

3 Loi des grands nombres

La loi des grands nombres “signifie” que l’espérance d’une variable aléatoire formalise la moyenne des résultats obtenus sur un grand nombre de répétitions de la même expérience. Les différents énoncés correspondent à des convergences plus ou moins fortes, au prix d’hypothèses plus ou moins coûteuses. On commence par deux versions faciles à démontrer.

Exercice 3.1 (Loi faible des grands nombres). Prouver le résultat suivant. Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. Si $\mathbb{E}[X_1^2] < \infty$ alors $\frac{1}{n} \sum_{k=1}^n X_k$ converge dans L^2 vers $\mathbb{E}[X_1]$.

Proof. $\mathbb{E}[(\bar{X}_n - \mathbb{E}[X_1])^2] = \frac{1}{n} \text{Var}(X_1) \rightarrow 0$. □

Exercice 3.2 (LGN L^4). *Prouver la loi forte des grands nombres dans le cas où les X_n ont un moment d'ordre 4.*

Proof. On peut supposer que les v.a. sont centrées. Ainsi,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}(X_1 + \dots, X_n)\right)^4\right] &= \frac{1}{n^4} \sum_{1 \leq i_1, \dots, i_4 \leq n} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] \\ &= \frac{1}{n^4} (n\mathbb{E}[X_1^4] + 3n(n-1)\mathbb{E}[X_1^2 X_2^2]) \\ &\leq \frac{C}{n^4}.\end{aligned}$$

Ainsi $\sum \mathbb{E}\left[\left(\frac{1}{n}(X_1 + \dots + X_n)\right)^4\right] < \infty$ et par Fubini-Tonelli, $\sum \left(\frac{1}{n}(X_1 + \dots + X_n)\right)^4 < \infty$ p.s. ce qui entraîne le résultat. $\square$

Notons que les hypothèses pour la loi faible peuvent être allégées ou changées. On énonce maintenant la loi forte des grands nombres, dont la démonstration est compliquée (cf par exemple Chabanol-Ruch).

Theorem 3.3 (Loi forte des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. Alors $\frac{1}{n} \sum_{k=1}^n X_k$ converge p.s. si et seulement si $\mathbb{E}[|X_1|] < +\infty$. Si l'une des deux conditions est vérifiée alors la limite est $\mathbb{E}[X_1]$.*

Un corollaire immédiat est le suivant, qui illustre par exemple le fait que si on répète un lancer de dé non pipé un grand nombre de fois, la fréquence d'apparition du 1 (ou d'une autre face), se rapproche de $\frac{1}{6}$.

Corollary 3.4. *Si (A_n) est une suite d'événements indépendants de même probabilité, alors $\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{A_k}$ converge p.s. vers $\mathbb{P}(A_1)$.*

Le théorème de Glivenko-Cantelli est une conséquence du corollaire précédent. Il est notamment utile pour illustrer une convergence en loi : il affirme que la fonction de répartition empirique d'un échantillon i.i.d. de loi donnée converge vers la fonction de répartition théorique de la loi en question. Rappelons la définition de la fonction de répartition empirique.

Definition 3.5. *Soit (X_n) une suite de v.a. i.i.d. La fonction de répartition empirique du vecteur aléatoire $(X_1, \dots, X_n)$ est la fonction (aléatoire) définie sur $\mathbb{R}$ par*

$$F_n(x) = \sum_{k=1}^n \mathbb{1}_{X_k \leq x}.$$

Theorem 3.6 (Théorème de Glivenko-Cantelli). *Soit (X_n) une suite de v.a. i.i.d. de même fonction de répartition F . Alors la fonction de répartition empirique F_n converge presque sûrement uniformément vers la fonction de répartition F :*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1.$$

Pour la preuve, on se réfère au Chabanol-Ruch par exemple. On verra dans la dernière section comment implémenter en pratique ce théorème pour illustrer une convergence en loi par ordinateur.

4 Théorème central limite

Theorem 4.1 (TCL 1D). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles i.i.d. dans L^2 . On note $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \text{Var}[X_1]$. Alors on a la convergence en loi*

$$\frac{1}{\sigma\sqrt{n}} \left(\sum_{k=1}^n X_k - n\mu \right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Proof. Une idée de développement, voir le Queffélec-Zuily par exemple... $\square$

Ce théorème suggère que la moyenne empirique des variables aléatoires converge vers son espérance (c'est la loi des grands nombres) à vitesse $O(\sqrt{n})$. Pour la preuve on peut se référer par exemple au Queffélec-Zuily.

A noter qu'une version plus faible de ce résultat, où les X_n sont des variables de Bernoulli, est enseignée en Terminale S : c'est le théorème de Moivre-Laplace.

Corollary 4.2. *Soit (S_n) une suite de v.a. binomiales de paramètres n, p . Alors*

$$\frac{1}{\sqrt{np(1-p)}}(S_n - np) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

en loi.

Exercice 4.3. *Démontrer le théorème de Moivre-Laplace directement.*

Proof. On peut supposer que les variables sont centrées. On pose $Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$. Par indépendance des X_k , la fonction caractéristique de Z_n est

$$\Phi_{Z_n}(u) = \Phi_{X_1}(u/\sqrt{n})^n.$$

Comme X_1 est L^2 , Φ_{X_1} est deux fois dérivable en 0, et

$$\Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right) = 1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n}\right)$$

puis en passant à la limite,

$$\lim_{n \rightarrow \infty} \Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right)^n = \exp\left(-\frac{\sigma^2 u^2}{2}\right).$$

$\square$

Un corollaire immédiat du TCL est le résultat suivant, qui permet d'obtenir des intervalles de confiance asymptotiques pour l'espérance, comme ce sera expliqué dans la partie statistiques.

Corollary 4.4. *Sous les hypothèses du TCL, pour $a > 0$,*

$$\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-a}^a \exp\left(-\frac{x^2}{2}\right) dx$$

où $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$.

La version multidimensionnelle fait intervenir les vecteurs Gaussiens.

Theorem 4.5 (TCL multidimensionnel). *Soit (X_n) une suite de vecteurs aléatoires de $\mathbb{R}^d$ i.i.d. dans L^2 . On note $\mu = \mathbb{E}[X_1]$, Γ la matrice de covariance, et $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors la suite $(\sqrt{n}(\bar{X}_n - \mu))_n$ converge en loi vers un vecteur Gaussien de loi $\mathcal{N}(0, \Gamma)$.*

Ce théorème est utile en statistiques notamment,

5 Illustration des théorèmes de convergence

5.1 Convergence presque sûre

Pour illustrer la convergence p.s. (loi des grands nombres), on simule un nombre suffisant de v.a. i.i.d., et on calcule à chaque fois la moyenne empirique, qu'on trace (en abscisses le nombre n de v.a. simulées et en ordonnées la valeur de $\frac{X_1 + \dots + X_n}{n}$). On effectue plusieurs simulations de la suite. Voici un exemple de code pour des v.a. uniformes sur $[0, 1]$ i.i.d.

```

# Paramètres
Nsimu = 5 # Nombre de simulations
n = 100 # Nombre de réalisations par simulations

# Réalisations des v.a. et calcul de la moyenne empirique
X = rd.random((Nsimu, n))
k = range(1,n+1)
X_bar = np.cumsum(X, axis=1)/k

plt.figure()

# Tracé graphique
plt.title("Loi des grands nombres")
plt.plot(k, X_bar.T)

# Une ligne pour la limite
plt.hlines(0.5, 1, n, label="Limite", color='k', linestyle='--')

# Des légendes
plt.xlabel("k")
plt.ylabel("X")

# Une grille en pointillés
plt.grid()

# Taille des axes
plt.axis([0, n, 0, 1]) # définit la taille des axes [xmin, xmax, ymin, ymax]

# Une légende avec les labels des tracés si besoin
plt.legend()

plt.show()

```

5.2 Convergence en loi

Il y a plusieurs façons d'illustrer la convergence en loi (TCL). A chaque fois, il s'agit de comparer l'échantillon obtenu sur un grand nombre de simulations avec le résultat théorique. Voici trois exemples différents.

5.2.1 Comparaison des fonctions de répartition

```

# Paramètres
Nsimu = 1000
n = 100

# Réalisations des v.a.
X = rd.random((Nsimu, n))
std = 1/np.sqrt(12)
X_bar = np.sum(X, axis=1)/n
Z = np.sqrt(n)*(X_bar-1/2)/std # Variables centrées réduites

# Calcul des fréquences cumulées croissantes
table = np.unique(Z, return_counts=True)

```

```

frequence_cum = np.cumsum(table[1])/Nsimu

# Tracé fdr empirique/fdr théorique
plt.figure()
plt.title("Théorème central limite")
plt.step(table[0], frequence_cum, label="Empirique")
plt.plot(table[0], st.norm.cdf(table[0]), label="Théorique", color='k', linestyle='--')
plt.xlabel("Z")
plt.ylabel("Probabilité")
plt.legend()
plt.show()

```

5.2.2 Histogramme

```

# Paramètres
Nsimu = 1000
n = 100

# Les réalisations
X = rd.random(Nsimu, n)
std = 1/np.sqrt(12)
X_bar = np.sum(X, axis=1)/n
Z = np.sqrt(n)*(X_bar-1/2)/std

# Valeurs en abscisses
x = np.linspace(min(Z), max(Z), num=1000)

# Tracé histogramme des valeurs/densité théorique
plt.figure()
plt.title("Théorème central limite")
plt.hist(Z, density=True, label="Empirique")
plt.plot(x, st.norm.pdf(x), label="Théorique", color='k', linestyle='--')
plt.xlabel("Z")
plt.ylabel("Probabilité")
plt.legend()
plt.show()

```

5.2.3 Diagramme en bâtons

```

# Paramètres
Nsimu = 10000
n = 10

# Réalisations de v.a. de Bernoulli
X = rd.binomial(n, 1/2, size=Nsimu)
table = np.unique(X, return_counts=True)
valeur = table[0]
frequence = table[1]/Nsimu

# Tracé fréquences empiriques/théoriques
plt.figure()
plt.title("Comparaison fréquences empiriques/théoriques")
plt.bar(valeur, frequence, label="Fréquence empirique") # Diagramme en bâtons

```

```
plt.vlines(valeur, [0]*(n+1), st.binom(n, 1/2).pmf(valeur),  
           color='black', linestyle='--', label="Probabilité théorique")  
plt.legend(loc="lower right")  
plt.xlabel("Position")  
plt.ylabel('Fréquence')  
plt.show()
```