

# Probabilité Agrégation

**Clément Pellegrini**

[clement.pellegrini@math.univ-toulouse.fr](mailto:clement.pellegrini@math.univ-toulouse.fr)

Institut de Mathématiques de Toulouse,  
Equipe de Statistique et Probabilité,  
Bureau 220 Bâtiment 1R1

- **1) Rappel de théorie de la mesure**
- **2) Variables aléatoires**
- **3) Espérance et Variance**
- **4) Transformées exponentielles**
- **5) Indépendance**
- **6) Conditionnement**

- **7) Convergence**
- **8) Loi des grands nombres et Théorème centrale limite**
- **9) Vecteurs gaussiens**
- **10) Estimateur du maximum de vraisemblance**

# **1) Rappel de théorie de la mesure**

# Définitions

Soit  $\Omega$  un ensemble

## Definition

$\mathcal{A} \subset \mathcal{P}(\Omega)$  est une tribu (ou  $\sigma$ -algèbre) sur  $\Omega$  si les conditions suivantes sont satisfaites

- 1  $\Omega \in \mathcal{A}$
- 2  $\mathcal{A}$  est stable par passage au complémentaire: si  $A \in \mathcal{A}$  alors  $A^c \in \mathcal{A}$
- 3  $\mathcal{A}$  est stable par réunion dénombrable: soit  $(A_n)_n$  une suite d'éléments de  $\mathcal{A}$  i.e  $A_n \in \mathcal{A}$  pour tout  $n \in \mathbb{N}$  alors  $\bigcup_n A_n \in \mathcal{A}$

- 1  $\{\emptyset, \Omega\}$  s'appelle la **tribu grossière**
- 2  $\mathcal{P}(\Omega)$  s'appelle la **tribu triviale**, c'est souvent celle considérée lorsque  $\Omega$  est discret ou dénombrable
- 3 Lorsque  $\Omega$  est un espace topologique c'est à dire muni d'une famille d'ouverts, la plus petite tribu contenant ces ouverts est appelée **tribu borélienne**. On la note  $\mathcal{B}(\Omega)$ . Pourquoi existe t'elle toujours?

# Définitions

Soit  $\Omega$  un ensemble

## Definition

$\mathcal{A} \subset \mathcal{P}(\Omega)$  est une tribu (ou  $\sigma$ -algèbre) sur  $\Omega$  si les conditions suivantes sont satisfaites

- 1  $\Omega \in \mathcal{A}$
- 2  $\mathcal{A}$  est stable par passage au complémentaire: si  $A \in \mathcal{A}$  alors  $A^c \in \mathcal{A}$
- 3  $\mathcal{A}$  est stable par réunion dénombrable: soit  $(A_n)_n$  une suite d'éléments de  $\mathcal{A}$  i.e  $A_n \in \mathcal{A}$  pour tout  $n \in \mathbb{N}$  alors  $\bigcup_n A_n \in \mathcal{A}$

- 1  $\{\emptyset, \Omega\}$  s'appelle la **tribu grossière**
- 2  $\mathcal{P}(\Omega)$  s'appelle la **tribu triviale**, c'est souvent celle considérée lorsque  $\Omega$  est discret ou dénombrable
- 3 Lorsque  $\Omega$  est un espace topologique c'est à dire muni d'une famille d'ouverts, la plus petite tribu contenant ces ouverts est appelée **tribu borélienne**. On la note  $\mathcal{B}(\Omega)$ . Pourquoi existe t'elle toujours?

## Définition

Soit  $X$  et  $Y$  deux ensembles munis de tribus  $\mathcal{A}$  pour  $X$  et  $\mathcal{B}$  pour  $Y$ . Une application  $f : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$  est dite mesurable si

$$\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A}$$

Rappel: de manière générale **toute application continue** entre espaces topologiques **est mesurable** si on munit ces espaces de leurs tribus boréliennes.

# Définitions

Un ensemble  $\Omega$  muni d'une tribu  $\mathcal{A}$  s'appelle un espace mesurable on le note  $(\Omega, \mathcal{A})$

## Definition

Une mesure  $\mu$  sur  $(\Omega, \mathcal{A})$  est une application de  $\mathcal{A} \rightarrow [0, +\infty]$  telle que

- 1  $\mu(\emptyset) = 0$
- 2 Si  $(A_n)_n$  est une suite dénombrable d'ensembles de  $\mathcal{A}$  deux à deux disjoints alors

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

- 1 Le triplet  $(\Omega, \mathcal{A}, \mu)$  est appelé espace mesuré.
- 2 Lorsque  $\mu$  est une mesure de masse 1 c'est à dire  $\mu(\Omega) = 1$  on parle de mesure de probabilité. Dans ce cas on note souvent la mesure  $\mu$  par  $\mathbb{P}$ .
- 3 Un espace de probabilité est donc un espace mesurable  $(\Omega, \mathcal{A})$  muni d'une mesure de probabilité  $\mathbb{P}$ .



# Définitions

Un ensemble  $\Omega$  muni d'une tribu  $\mathcal{A}$  s'appelle un espace mesurable on le note  $(\Omega, \mathcal{A})$

## Definition

Une mesure  $\mu$  sur  $(\Omega, \mathcal{A})$  est une application de  $\mathcal{A} \rightarrow [0, +\infty]$  telle que

- 1  $\mu(\emptyset) = 0$
- 2 Si  $(A_n)_n$  est une suite dénombrable d'ensembles de  $\mathcal{A}$  deux à deux disjoints alors

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

- 1 **Le triplet  $(\Omega, \mathcal{A}, \mu)$  est appelé espace mesuré.**
- 2 **Lorsque  $\mu$  est une mesure de masse 1** c'est à dire  $\mu(\Omega) = 1$  on parle de **mesure de probabilité**. Dans ce cas on note souvent la mesure  $\mu$  par  $\mathbb{P}$ .
- 3 **Un espace de probabilité** est donc un espace mesurable  $(\Omega, \mathcal{A})$  muni d'une mesure de probabilité  $\mathbb{P}$ .

On rappelle que

①  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

②  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

③ Si  $A \subset B$ :  $\mathbb{P}(A) \leq \mathbb{P}(B)$

④ Si  $(A_n)$  est une suite croissante i.e  $A_n \subset A_{n+1}$

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_n A_n\right)$$

⑤ Si  $(A_n)$  est une suite décroissante i.e  $A_{n+1} \subset A_n$

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_n A_n\right)$$

## **2) Variables aléatoires**

# Définitions

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité

## Definition

On appelle **variable aléatoire (v.a) toute application mesurable**  $X$  d'un espace de probabilité  $\Omega$  dans  $\mathbb{R}$  munit de la tribu borélienne  $\mathcal{B}(\mathbb{R})$ .

Si la **variable aléatoire est à valeurs dans  $\mathbb{R}^d$  on parle de vecteur aléatoire** (couple si  $d=2$ ).

## Definition

Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a, on appelle loi de  $X$  la mesure de probabilité  $\mathbb{P}_X$  qui est la mesure image sur  $\mathbb{R}$  de  $\mathbb{P}$  par  $X$  c'est à dire

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) := \mathbb{P}(X \in A)$$

Pour parler de la loi d'une v.a  $X$  on utilise la **notation  $\sim$** . Par exemple  $X \sim \mathcal{P}(\lambda)$  pour une loi de Poisson de paramètre  $\lambda$ .

# Définitions

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité

## Definition

On appelle **variable aléatoire (v.a) toute application mesurable**  $X$  d'un espace de probabilité  $\Omega$  dans  $\mathbb{R}$  munit de la tribu borélienne  $\mathcal{B}(\mathbb{R})$ .

Si la **variable aléatoire est à valeurs dans  $\mathbb{R}^d$  on parle de vecteur aléatoire** (couple si  $d=2$ ).

## Definition

Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a, on appelle loi de  $X$  la mesure de probabilité  $\mathbb{P}_X$  qui est la mesure image sur  $\mathbb{R}$  de  $\mathbb{P}$  par  $X$  c'est à dire

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) := \mathbb{P}(X \in A)$$

Pour parler de la loi d'une v.a  $X$  on utilise la **notation  $\sim$** . Par exemple  $X \sim \mathcal{P}(\lambda)$  pour une loi de Poisson de paramètre  $\lambda$ .

# Définitions

Soit  $X$  une v.a.

## Definition

- On appelle atome de  $X$  (ou de sa loi) tout  $a \in \mathbb{R}$  tel que  $\mathbb{P}(X = a) \neq 0$
- On appelle **support (topologique)** d'une v.a  $X$  le plus petit fermé  $F$  tel que  $\mathbb{P}(X \in F) = 1$ . On notera  $S(X)$  le support d'une v.a  $X$ .

## Definition

La tribu engendrée par  $X$ , notée  $\sigma(X)$ , est la plus petite tribu  $\mathcal{G}$  sur  $\Omega$  qui rend l'application  $X : (\Omega, \mathcal{G}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mesurable.

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$$

- 1 On interprète  $\sigma(X)$  comme l'ensemble de tous les évènements accessibles à partir de la connaissance de  $X$ . Cela représente en quelque sorte toute l'information nécessaire véhiculée par  $X$ .
- 2 On généralise cela à plusieurs v.a  $X_1, \dots, X_n$  et on note  $\sigma(X_1, \dots, X_n)$  la plus petite tribu rendant les  $X_i$  mesurables.

# Définitions

Soit  $X$  une v.a.

## Definition

- On appelle atome de  $X$  (ou de sa loi) tout  $a \in \mathbb{R}$  tel que  $\mathbb{P}(X = a) \neq 0$
- On appelle **support (topologique)** d'une v.a  $X$  **le plus petit fermé  $F$  tel que  $\mathbb{P}(X \in F) = 1$** . On notera  $S(X)$  le support d'une v.a  $X$ .

## Definition

La tribu engendrée par  $X$ , notée  $\sigma(X)$ , est la plus petite tribu  $\mathcal{G}$  sur  $\Omega$  qui rend l'application  $X : (\Omega, \mathcal{G}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mesurable.

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$$

- 1 On interprète  $\sigma(X)$  comme l'ensemble de tous les évènements accessibles à partir de la connaissance de  $X$ . Cela représente en quelque sorte toute l'information nécessaire véhiculée par  $X$ .
- 2 On généralise cela à plusieurs v.a  $X_1, \dots, X_n$  et on note  $\sigma(X_1, \dots, X_n)$  la plus petite tribu rendant les  $X_j$  mesurables.

## Definition

On appelle fonction de répartition d'une v.a  $X : \Omega \rightarrow \mathbb{R}$  la fonction  $F_X$  définie sur  $\mathbb{R}$  par

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}_X(] - \infty, t])$$

## Proposition

- 1  $F_X$  est croissante
- 2  $\lim_{t \rightarrow -\infty} F_X(t) = 0$  et  $\lim_{t \rightarrow +\infty} F_X(t) = 1$
- 3  $F_X$  est continue à droite et admet une limite à gauche
- 4 Si  $X$  n'a pas d'atome  $F_X$  est continue



## Definition

On appelle fonction de répartition d'une v.a  $X : \Omega \rightarrow \mathbb{R}$  la fonction  $F_X$  définie sur  $\mathbb{R}$  par

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}_X(] - \infty, t])$$

## Proposition

- 1  $F_X$  est croissante
- 2  $\lim_{t \rightarrow -\infty} F_X(t) = 0$  et  $\lim_{t \rightarrow +\infty} F_X(t) = 1$
- 3  $F_X$  est continue à droite et admet une limite à gauche
- 4 Si  $X$  n'a pas d'atome  $F_X$  est continue

- Les éventuels sauts de  $F_X$  qui sont en nombre dénombrable (pourquoi?) correspondent aux atomes de  $X$
- Le quantile d'ordre  $\alpha$  est toute valeur  $z_\alpha$  telle que

$$\mathbb{P}(X \leq z_\alpha) = \alpha$$

ou

$$\mathbb{P}(X > z_\alpha) = 1 - \alpha$$

- Médiane: tout quantile d'ordre  $1/2$
- Inverse de la fonction de répartition. Pour  $u \in ]0, 1[$ , on pose

$$G(u) = \inf\{t, |F(t) \geq u\}$$

On a que  $G(u) \leq t \Leftrightarrow F(t) \geq u$ , cela donne une méthode de simulation.

## Definition

**Une variable aléatoire  $X$  est dite discrète** si elle est à valeurs dans un ensemble discret ou dénombrable. En particulier  $S(X)$  est fini ou dénombrable.

On pose  $S(X) = \{a_i, i \in I\}$  où  $I$  est un ensemble fini ou dénombrable. Si on pose

$$p_i = \mathbb{P}(X = a_i), i \in I$$

La loi de  $X$  est

$$P_X = \sum_{i \in I} p_i \delta_{a_i}$$

Rq: Si  $(X, Y)$  est un couple de v.a discrète on obtient la loi de  $X$  par la formule

$$P_X(x) = \sum_{y \in S(Y)} \mathbb{P}_{(X,Y)}(x, y), \forall x \in S(X)$$

# Exemples usuels

- Les v.a constantes. Si  $X = c$  presque sûrement alors  $S(X) = \{c\}$  et

$$\mathbb{P}_X = \delta_c, \quad \mathbb{P}(X = c) = 1$$

- Si  $S(X) = \{a_1, \dots, a_n\}$  et  $X$  prend la valeur  $a_i$  avec probabilité  $p_i$  on a

$$\mathbb{P}_X = \sum_{i=1}^n p_i \delta_{a_i}, \quad \mathbb{P}(X = a_i) = p_i$$

On dit que  $X$  suit la loi uniforme (ou équirépartie) sur  $\{a_1, \dots, a_n\}$  si  $p_i = \frac{1}{n}$  pour tout  $i \in \{1, \dots, n\}$  dans ce cas

$$\mathbb{P}_X = \sum_{i=1}^n \frac{1}{n} \delta_{a_i}, \quad \mathbb{P}(X = a_i) = \frac{1}{n}$$

(exemple du dé à six face)

# Exemples usuels

- Les v.a constantes. Si  $X = c$  presque sûrement alors  $S(X) = \{c\}$  et

$$\mathbb{P}_X = \delta_c, \quad \mathbb{P}(X = c) = 1$$

- Si  $S(X) = \{a_1, \dots, a_n\}$  et  $X$  prend la valeur  $a_i$  avec probabilité  $p_i$  on a

$$\mathbb{P}_X = \sum_{i=1}^n p_i \delta_{a_i}, \quad \mathbb{P}(X = a_i) = p_i$$

On dit que  $X$  suit la loi uniforme (ou équirépartie) sur  $\{a_1, \dots, a_n\}$  si  $p_i = \frac{1}{n}$  pour tout  $i \in \{1, \dots, n\}$  dans ce cas

$$\mathbb{P}_X = \sum_{i=1}^n \frac{1}{n} \delta_{a_i}, \quad \mathbb{P}(X = a_i) = \frac{1}{n}$$

(exemple du dé à six face)

# Exemples usuels

- Si  $X$  modélise un succès (obtenir 1) ou un échec (obtenir 0). Son support est  $S(X) = \{0, 1\}$  et sa loi est

$$\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0, \quad \mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$$

où  $p$  est la probabilité de succès. Il s'agit de la loi de **Bernoulli**  $X \sim \mathcal{B}(p)$

- Si la v.a  $X$  modélise le nombre de succès dans une suite de  $n$  expériences identiques et indépendantes avec chacune une probabilité  $p$  de succès alors  $S(X) = \{0, \dots, n\}$  et on a

$$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Dans ce cas on dit que  $X$  suit une loi **binomiale de paramètre**  $\underline{p}$ , la probabilité de succès, et  $\underline{n}$ , le nombre d'expérience, :  $X \sim \mathcal{B}(n, p)$ .

- Si  $X$  modélise un succès (obtenir 1) ou un échec (obtenir 0). Son support est  $S(X) = \{0, 1\}$  et sa loi est

$$\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0, \quad \mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$$

où  $p$  est la probabilité de succès. Il s'agit de la loi de **Bernoulli**  $X \sim \mathcal{B}(p)$

- Si la v.a  $X$  modélise le nombre de succès dans une suite de  $n$  expériences identiques et indépendantes avec chacune une probabilité  $p$  de succès alors  $S(X) = \{0, \dots, n\}$  et on a

$$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Dans ce cas on dit que  $X$  suit une loi **binomiale de paramètre**  $\underline{p}$ , la probabilité de succès, et  $\underline{n}$ , le nombre d'expérience, :  $X \sim \mathcal{B}(n, p)$ .

- La variable aléatoire  $X$  qui modélise le rang du premier succès dans une suite d'expériences indépendantes et identiques a pour support  $S(X) = \mathbb{N}^*$  et on a

$$\mathbb{P}_X = \sum_{k=1}^{\infty} (1-p)^{k-1} p \delta_k, \quad \mathbb{P}(X = k) = (1-p)^{k-1} p$$

Dans ce cas on dit que  $X$  suit une **loi géométrique** de paramètre  $\underline{p}$ :  
 $X \sim \mathcal{G}(p)$

- On dit que la variable aléatoire  $X$  suit une **loi de Poisson** de paramètre  $\lambda > 0$ :  $X \sim \mathcal{P}(\lambda)$  si  $S(X) = \mathbb{N}$  et

$$\mathbb{P}_X = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k, \quad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Une telle loi sert à modéliser des petits effectifs



- La variable aléatoire  $X$  qui modélise le rang du premier succès dans une suite d'expériences indépendantes et identiques a pour support  $S(X) = \mathbb{N}^*$  et on a

$$\mathbb{P}_X = \sum_{k=1}^{\infty} (1-p)^{k-1} p \delta_k, \quad \mathbb{P}(X = k) = (1-p)^{k-1} p$$

Dans ce cas on dit que  $X$  suit une **loi géométrique** de paramètre  $\underline{p}$ :  
 $X \sim \mathcal{G}(p)$

- On dit que la variable aléatoire  $X$  suit une **loi de Poisson** de paramètre  $\underline{\lambda > 0}$ :  $X \sim \mathcal{P}(\lambda)$  si  $S(X) = \mathbb{N}$  et

$$\mathbb{P}_X = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k, \quad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Une telle loi sert à modéliser des petits effectifs

## Proposition

*Soit  $X$  une v.a discrète de support  $S(X) = \{a_i, i \in I\}$  avec les probabilités  $p_i, i \in I$  associées. La fonction de répartition est constante par morceaux avec des sauts de taille  $p_i$  en  $a_i$*

$$F_X(t) = \sum_{j \leq i} p_j, \quad \text{si } t \in [a_i, a_{i+1}[ , i \in I$$

## Definition

Une v.a  $X$  est dite à densité  $f$  si pour tout borélien  $A$

$$\mathbb{P}(X \in A) = \int_A f(x) dx$$

- Dans ce cas sa loi  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue et sa **dérivée de Radon Nykodim** est  $f$
- On a nécessairement que  $f \geq 0$
- $\int_{\mathbb{R}} f(x) dx = 1$
- Pour un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  on a pour tout borélien  $A$  de  $\mathbb{R}^d$ :

$$\mathbb{P}_X(A) = \int_A f(x_1, \dots, x_d) dx_1 \dots dx_d$$

## Definition

Une v.a  $X$  est dite à densité  $f$  si pour tout borélien  $A$

$$\mathbb{P}(X \in A) = \int_A f(x) dx$$

- Dans ce cas sa loi  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue et sa **dérivée de Radon Nykodim** est  $f$
- On a nécessairement que  $f \geq 0$
- $\int_{\mathbb{R}} f(x) dx = 1$
- Pour un vecteur aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  on a pour tout borélien  $A$  de  $\mathbb{R}^d$ :

$$\mathbb{P}_X(A) = \int_A f(x_1, \dots, x_d) dx_1 \dots dx_d$$

- Loi uniforme sur  $[0, 1]$ :  $X \sim \mathcal{U}([0, 1])$  la densité

$$f(x) = \mathbf{1}_{[0,1]}(x)$$

- Loi uniforme sur  $[a, b]$ :  $X \sim \mathcal{U}([a, b])$  la densité

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$$

- Loi exponentielle de paramètre  $\lambda > 0$ :  $X \sim \mathcal{E}(\lambda)$  est de densité

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}^+}(x)$$

# Exemples usuels

- Loi normale centrée réduite  $X \sim \mathcal{N}(0, 1)$  est de densité

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

- Loi normale de moyenne  $m$  et de variance  $\sigma^2$ :  $X \sim \mathcal{N}(m, \sigma^2)$  est de densité

$$f(x) = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

- Loi de Cauchy de paramètre  $a > 0$   $X \sim C(a)$  est de densité

$$f(x) = \frac{a}{\pi(a^2 + x^2)}$$

- Loi Gamma de paramètres  $n \in \mathbb{N}^*$ ,  $\lambda > 0$ :  $X \sim \gamma(n, \lambda)$  est de densité

$$f(x) = \frac{x^{n-1} \lambda^n e^{-\lambda x}}{\Gamma(n)} \mathbf{1}_{\mathbb{R}^+}(x)$$

## Proposition

Si  $X$  est une v.a de densité  $f$ , sa fonction de répartition  $F_X$  vérifie

- 1  $\forall t \in \mathbb{R}, F_X(t) = \int_{-\infty}^t f(x)dx$
- 2  $F_X$  est continue sur  $\mathbb{R}$
- 3 Si  $f$  est continue au point  $t_0$ , alors  $F_X$  est dérivable en  $t_0$  de dérivée  $F'_X(t_0) = f(t_0)$

- Rq:  $\mathbb{P}(X = a) = \int_a^a f(x)dx = 0$ . Cela implique que dans le calcul des probabilités, on peut considérer indifféremment des inégalités larges ou strictes.
- Si  $(X, Y)$  est un couple de densité  $f(x, y)$ , alors  $X$  est à densité et celle-ci s'obtient par

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)dy.$$

# 3) **Espérance et Variance**



# Formule du transfert

Soit  $X$  une v.a (ou vecteur), on rappelle que la loi de  $X$  est décrite à travers la formule  $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$ , pour tout  $A$  borélien. On pose alors

$$\mathbb{E}[\mathbf{1}_A(X)] = \mathbb{P}_X(A) = \mathbb{E}[\phi(X)],$$

avec  $\phi(x) = \mathbf{1}_A(x)$ . On généralise cette formule pour toute fonction  $\phi$  continue et bornée

## Theorem

Soit  $X$  une v.a, pour toute fonction continue bornée  $\phi$

$$\mathbb{E}[\phi(X)] = \int \phi(X(\omega))d\mathbb{P}(\omega) = \int \phi(x)d\mathbb{P}_X(x)$$

- Dans le cas discret  $\mathbb{E}[\phi(X)] = \sum_{x \in \mathcal{S}(X)} \phi(x)\mathbb{P}[X = x]$
- Dans le cas continu  $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x)f(x)dx$

# Formule du transfert

- Rq:  $X$  a pour densité  $f$  ssi  $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x)f(x)dx$  pour tout  $\phi$  continue bornée
- Soit  $X$  une v.a à densité  $f$  et  $g$  une fonction bijective de  $\mathbb{R}$  dans  $\mathbb{R}$  de classe  $C^1$ . On pose  $Y = g(X)$ . Alors la densité de  $Y$  est

$$f_Y(y) = f_X(g^{-1}(y))|(g^{-1})'(y)|$$

- Si  $z \sim \mathcal{U}([-1, 1])$  alors  $|Z| \sim \mathcal{U}([0, 1])$
- Si  $X \sim \mathcal{N}(m, \sigma^2)$  alors

$$\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1)$$

- Si  $X \sim \mathcal{N}(0, 1)$  quelle est la loi de  $Y = X^2$
- Pour des vecteurs aléatoires on utilise le même type de calcul mais il faut utiliser un changement de variable en dimension supérieure.

## Definition

Une v.a  $X$  à un moment d'ordre  $p \geq 1$  si

$$\mathbb{E}(|X|^p) < +\infty$$

Si  $X$  a un moment d'ordre 1 on appelle espérance la quantité

$$\mathbb{E}[X]$$

Si  $X$  a un moment d'ordre 2 on appelle **variance** la quantité

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

On appelle **écart type** de  $X$  la quantité

$$\sigma_X = \sqrt{\text{Var}(X)}$$

- Inégalité de Markov

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}(|X|)}{t}$$

- Inégalité de Bienaymé-Tchebychev

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

- Comme l'espace est de masse finie si une v.a  $X$  admet un moment d'ordre  $p \geq 1$  alors elle admet un moment d'ordre  $q$  pour tout  $q$  tel que  $p \geq q \geq 1$ .
- L'espérance est linéaire.
- $\text{Var}(aX) = a^2 \text{Var}(X)$ .
- $\text{Var}(X) = 0$  ssi  $X$  est constante
- L'espace vectoriel  $L^p(\Omega, \mathcal{A}, \mathbb{P}) = \{X \text{ v.a}, \mathbb{E}(|X|^p) < +\infty\}$  est **un espace de Banach**.

# Exemples

- $X \sim \mathcal{B}(p)$  alors  $\mathbb{E}[X] = p$ ,  $\text{Var}(X) = p(1 - p)$
- $X \sim \mathcal{B}(n, p)$  alors  $\mathbb{E}[X] = np$ ,  $\text{Var}(X) = np(1 - p)$
- $X \sim \mathcal{G}(p)$  alors  $\mathbb{E}[X] = \frac{1}{p}$ ,  $\text{Var}(X) = \frac{1-p}{p^2}$
- $X \sim \mathcal{P}(\lambda)$  alors  $\mathbb{E}[X] = \lambda$ ,  $\text{Var}(X) = \lambda$
- $X \sim \mathcal{U}([a, b])$  alors  $\mathbb{E}[X] = \frac{b+a}{2}$ ,  $\text{Var}(X) = \frac{(a-b)^2}{12}$
- $X \sim E(\lambda)$  alors  $\mathbb{E}[X] = \frac{1}{\lambda}$ ,  $\text{Var}(X) = \frac{1}{\lambda^2}$
- $X \sim C(a)$  alors  $\mathbb{E}[|X|] = +\infty$
- $X \sim \mathcal{N}(m, \sigma^2)$  alors  $\mathbb{E}(X) = m$ ,  $\text{Var}(X) = \sigma^2$

# 4) Transformée exponentielle

## Definition

**La fonction caractéristique** d'une v.a  $X$  est la fonction définie par

$$\phi_X(t) = \mathbb{E}[e^{itX}], \forall t \in \mathbb{R}$$

**La fonction caractéristique d'un vecteur aléatoire**  $X$  à valeurs dans  $\mathbb{R}^d$  est la fonction définie par

$$\phi_X(u) = \mathbb{E}[e^{i\langle u, X \rangle}], \forall u \in \mathbb{R}^d,$$

où  $\langle, \rangle$  désigne le produit scalaire euclidien sur  $\mathbb{R}^d$ .

La fonction caractéristique est continue bornée par 1 et vaut 1 en 0.

# Fonction caractéristique

- $X \sim \mathcal{B}(p)$  alors  $\phi_X(t) = 1 - p + pe^{it}$
- $X \sim \mathcal{B}(n, p)$  alors  $\phi_X(t) = (1 - p + pe^{it})^n$
- $X \sim \mathcal{P}(\lambda)$  alors  $\phi_X(t) = \exp(\lambda(e^{it} - 1))$
- $X \sim \mathcal{U}([a, b])$  alors  $\phi_X(t) = \frac{e^{ibt} - e^{iat}}{(b-a)it}$
- $X \sim \mathcal{E}(\lambda)$  alors  $\phi_X(t) = \frac{\lambda}{\lambda - it}$
- $X \sim \mathcal{C}(a)$  alors  $\phi_X(t) = \exp(-a|t|)$
- $X \sim \mathcal{N}(m, \sigma^2)$  alors  $\phi_X(t) = \exp(imt - \frac{\sigma^2 t^2}{2})$



## Proposition

*Si une v.a  $X$  admet un moment d'ordre  $p$  alors sa fonction caractéristique est dérivable  $p$  fois et on a*

$$\phi_X^{(p)}(0) = i^p \mathbb{E}[X^p]$$

*Une réciproque partielle est si  $\phi_X$  est  $p \geq 2$  fois dérivable en 0 alors  $X$  admet des moments d'ordre  $2[p/2]$*

- Fonction génératrice des moments pour une v.a  $X$  tel que  $S(X) \subset \mathbb{N}$  et  $p_k = \mathbb{P}(X = k)$

$$G_X(t) = \mathbb{E}[t^X] = \sum_k p_k t^k$$

Cette fonction est  $C^\infty$  sur  $[0, 1[$  et dérivable à l'ordre  $p$  en 1 si  $\mathbb{E}[X^p] < +\infty$

$$G_X^{(k)}(0) = k! p_k, k \in \mathbb{N}$$

De plus si le moment d'ordre 1 existe  $G'_X(1) = \mathbb{E}(X)$

- Transformée de Laplace. Pour une v.a  $X$ , on appelle transformée de Laplace de  $X$

$$\phi_X(t) = \mathbb{E}[e^{tX}]$$

# 5) Indépendance

## Definition

- Deux évènements  $A$  et  $B$  d'une tribu  $\mathcal{A}$  sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

- Deux variables aléatoires sont dites indépendantes si pour tous boréliens  $A$  et  $B$  on a

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

On note parfois  $X \perp\!\!\!\perp Y$

- Deux tribus  $\mathcal{F}$  et  $\mathcal{G}$  sont dites indépendantes si pour tous événements  $F \in \mathcal{F}$  et  $G \in \mathcal{G}$ ,  $F$  et  $G$  sont indépendants.

On voit que  $X \perp\!\!\!\perp Y$  ssi  $\sigma(X) \perp\!\!\!\perp \sigma(Y)$

## Proposition

Deux v.a  $X$  et  $Y$  sont indépendantes ssi la loi  $\mathbb{P}_{X,Y}$  est la loi produit des marginales  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  i.e

$$\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y$$

- Dans le cas discret cela signifie que pour tout  $x \in S(X)$  et pour tout  $y \in S(Y)$  on a

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

- Dans le cas continu

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

## Theorem

Deux v.a  $X$  et  $Y$  sont indépendantes ssi

$$\phi_{X,Y}(s, t) = \phi_X(s)\phi_Y(t)$$

pour tout  $s, t \in \mathbb{R}$ .

## Proposition

Si  $X$  et  $Y$  sont deux v.a indépendantes alors la loi de  $X + Y$  est donnée par  $\mathbb{P}_X * \mathbb{P}_Y$ .

En particulier si  $X$  et  $Y$  ont pour densité  $f$  et  $g$  on a

$$f_{X+Y}(x) = \int f(y)g(x - y)dy$$

On a également

$$\phi_{X+Y}(\cdot) = \phi_X(\cdot)\phi_Y(\cdot)$$

- Si  $X \sim \mathcal{P}(\lambda)$  et  $Y \sim \mathcal{P}(\mu)$  deux Poissons indépendantes alors  $X + Y \sim \mathcal{P}(\lambda + \mu)$
- Soit  $X$  et  $Y$  deux v.a normales centrées réduites indépendantes alors  $X + Y$  et  $X - Y$  sont encore des v.a normales indépendantes.

## Definition

Une suite de v.a  $(X_n)$  est dite indépendante si pour tout  $i_1, \dots, i_k, i_1 < i_2 < \dots < i_k$  et pour tout  $k \in \mathbb{N}$

$$\mathbb{P}(X_{i_1} \in A_1, \dots, X_{i_k} \in A_k) = \mathbb{P}(X_{i_1} \in A_1) \dots \mathbb{P}(X_{i_k} \in A_k),$$

pour tout  $A_1, \dots, A_k$  dans la tribu.

En particulier pour le cas d'une famille finie  $X_1, \dots, X_N$ , cette famille est dite indépendante si:

$$\mathbb{P}(X_1 \in A_1, \dots, X_N \in A_N) = \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_N \in A_N),$$

pour tout  $A_1, \dots, A_N$  dans la tribu

- Soit  $X_1, \dots, X_n$  une famille de v.a.i.i.d de loi de  $\mathcal{B}(p)$  alors

$$X_1 + \dots + X_n \sim \mathcal{B}(n, p)$$

## **6) Conditionnement**



## Definition

Soit  $B$  un évènement de probabilité non nulle  $\mathbb{P}(B) \neq 0$ . Pour tout évènement  $A$  on définit la probabilité conditionnelle de  $A$  sachant  $B$  par

$$\mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$
- L'application  $\mathbb{P}(\cdot|B)$  définit une mesure  $(\Omega, \mathcal{A})$
- Si  $A \perp B$  alors  $\mathbb{P}(A|B) = \mathbb{P}(A)$

# Formule des probas totales et formule de Bayes

- Probabilité totale:

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

- Deux joueurs  $A$  et  $B$  avec respectivement une fortune  $a$  et  $b$  joue à pile ou face avec où pile arrive avec probabilité  $p$ . Le joueur  $A$  gagne 1 euro que lui donne  $B$  si la pièce tombe sur pile et perd 1 euro qu'il donne à  $B$  si la pièce tombe sur face. On pose  $u_a$  la probabilité que  $A$  soit ruiné. On a

$$u_a = pu_{a+1} + (1 - p)u_{a-1}$$

- Formule de Bayes:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Dans la pratique on utilise souvent la formule des probabilités totales pour calculer  $\mathbb{P}(A)$

## Proposition

Soit  $A_1, \dots, A_N$  une partition de  $\Omega$  alors

$$\mathbb{P}(A) = \sum_{i=1}^N \mathbb{P}(A|A_i)\mathbb{P}(A_i)$$

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i)\mathbb{P}(A_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^N \mathbb{P}(A|A_i)\mathbb{P}(A_i)}$$

- Soient  $X$  et  $Y$  deux variables aléatoires. Le théorème de désintégration des mesures nous dit que

$$\mathbb{P}(Y \in A, X \in B) = \int \mathbb{P}(Y \in A | X = x) \mathbb{P}_X(dx) = \mathbb{E}[\mathbf{1}_B \mathbb{P}(Y \in A | X)]$$

- Attention la quantité  $\mathbb{P}(Y \in A | X = x)$  est une notation cela correspond à la densité de Radon Nykodym
- La famille  $(\mathbb{P}(Y \in \cdot | X = x))_{x \in \mathbb{R}}$  s'appelle la famille des probabilités conditionnelles de  $Y$  sachant  $X$ .
- La loi conditionnelle de  $Y$  sachant  $X$  est noté  $\mathbb{P}(Y \in \cdot | X)$

- Dans le cas discret on obtient facilement les probabilités conditionnelles. En particulier

$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}$$

On a alors

$$\mathbb{P}(Y \in \cdot | X) = \sum_{x \in \mathcal{S}(X)} \mathbb{P}(Y = y | X = x) \mathbf{1}_{X=x}$$

- Dans le cas continu on parle de densité conditionnelle et on pose

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} \mathbf{1}_{f_X(x) > 0}$$

avec

$$f_X(x) = \int f(x, y) dy$$

# 7) Convergence stochastique

## Definition

Soit  $(X_n)$  une suite de v.a et  $X$  une v.a. On dit que  $(X_n)$  converge vers  $X$

- **Presque sûrement p.s** si  $\mathbb{P}(\lim X_n = X) = 1$

on note

$$X_n \xrightarrow{p.s} X$$

- **En norme  $L^p$**  si  $\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|^p] = 0$

on note

$$X_n \xrightarrow{L^p} X$$

- **En probabilité** si  $\forall \epsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0$

on note

$$X_n \xrightarrow{\mathbb{P}} X$$

- **En loi** si pour toute fonction continue bornée  $f$  on a

$$\lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad \text{on note} \quad X_n \xrightarrow{\text{Loi}} X$$

# Convergence en loi

On suppose  $(X_n)$  à valeurs dans un espace métrique

## Theorem

Les assertions suivantes sont équivalentes

- $(X_n)$  converge en loi vers  $X$
- pour toute fonction  $\varphi$  bornée et uniformément continue sur  $E$ ,

$$\lim_n \mathbb{E} [\varphi(X_n)] = \mathbb{E} [\varphi(X)]$$

- Pour tout fermé  $F$  on a  $\limsup_n \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$
- Pour tout ouvert  $O$  on a  $\liminf_n \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$
- Pour tout borélien  $A$  dont la frontière  $\partial A$  vérifie  $\mathbb{P}(X \in \partial A) = 0$  on a

$$\lim_n \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A).$$



# Convergence en loi

Pour une v.a on note  $F_X$  sa fonction de répartition et  $\phi_X$  sa fonction caractéristique

## Theorem

$(X_n)$  converge en loi vers  $X$  si et seulement si

$$F_{X_n}(t) \rightarrow F_X(t)$$

en tout point de continuité de  $F_X$  i.e en tout  $t$  tel que  $\mathbb{P}(X = t) = 0$

## Theorem

$(X_n)$  converge en loi vers  $X$  si et seulement si

$$\phi_{X_n}(t) \rightarrow \phi_X(t)$$

pour tout  $t \in \mathbb{R}$ .

Rappel des convergences classiques en termes probabilistes

## Theorem

- Soit  $(X_n)$  une suite croissante de v.a positives et  $\lim_n X_n = X$  alors

$$\lim_n \mathbb{E}[X_n] = \mathbb{E}[X]$$

- Soit  $(X_n)$  une suite croissante de v.a positives

$$\mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n]$$

- Soit  $(X_n)$  une suite de v.a telle que  $X_n$  converge p.s vers  $X$ . Soit  $Y$  tel que  $\mathbb{E}[|Y|] < \infty$  et  $|X_n| < |Y|$  alors

$$\lim_n \mathbb{E}[X_n] = \mathbb{E}[\lim_n X_n]$$

On a les liens suivants

- Convergence p.s  $\implies$  Convergence en probabilité
- Convergence  $L^p \implies$  Convergence  $L^1 \implies$  Convergence en probabilité
- Toute Convergence  $\implies$  Convergence en loi
- Convergence p.s + domination  $\implies$  Convergence  $L^1$
- Convergence  $L^1 \implies$  Convergence p.s pour une sous-suite

Montrer des convergences presque sûres? Le Lemme de Borel Canteli

## Lemma

Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'évènements

- Si  $\sum \mathbb{P}(A_n) < +\infty$  alors

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) := \mathbb{P}\left(\bigcap_{n \geq 0} \bigcup_{k \geq n} A_k\right) = 0$$

En d'autres termes seuls un nombre fini d'évènements  $A_n$  sont réalisés.

- Si  $\sum \mathbb{P}(A_n) = +\infty$  et les  $A_n$  sont indépendants alors

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) := \mathbb{P}\left(\bigcap_{n \geq 0} \bigcup_{k \geq n} A_k\right) = 1$$

En d'autres termes une infinité d'évènements  $A_n$  sont réalisés.

## Proposition

*$(X_n)$  converge en probabilité vers  $X$  si et seulement si de toute suite extraite on peut extraire une sous suite qui converge presque sûrement vers  $X$*

- La distance  $d(X, Y) = \mathbb{E}[|X - Y| \wedge 1]$  métrise la convergence en probabilité
- La topologie de la convergence presque sûre n'est pas métrisable

## Des exemples et des contre-exemples

- Pour tout  $n$  on considère la variable aléatoire  $X_n$  qui prend les valeurs  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$  avec la probabilité uniforme alors  $(X_n)$  converge en loi vers  $X \sim \mathcal{U}([0, 1])$ .
- Une suite de variables aléatoires  $(X_n)$  telle que  $X_n \sim \mathcal{B}(p_n)$  pour tout  $n \in \mathbb{N}$  (fournit des exemples et des contre-exemples...)
- Soit  $(X_n)$  une suite de v.a.i.i.d telle  $X_0 \sim \mathcal{B}(\frac{1}{2})$ , alors

$$\sum_{n=0}^{\infty} \frac{X_n}{2^n}$$

converge presque sûrement vers une v.a  $X \sim \mathcal{U}([0, 1])$ . Réciproque?

- $(X_n)$  telle que  $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$  et  $\mathbb{P}[X_n = a_n] = \frac{1}{n}$

- $(X_n)$  converge en loi vers  $X$  et  $(Y_n)$  converge en loi vers  $Y$  n'implique pas que le couple converge en loi vers  $(X, Y)$ .

## Proposition

*$(X_n)$  converge en loi vers une constante  $c$  alors  $(X_n)$  converge en probabilité vers  $c$*

- Si  $(X_n)$  converge en loi vers  $X$  et  $(Y_n)$  converge en loi vers  $c$  alors  $(X_n, Y_n)$  converge en loi vers  $(X, c)$  (Slutsky)

# **8) Loi des grands nombres et Théorème centrale limite**



Le but de cette partie est de décrire le comportement de somme de v.a qui la plupart du temps seront i.i.d.

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m)$$

## Theorem

Soit  $(X_n)$  une suite de v.a.i.i.d et  $L^2$  alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_0]$$

- Soit  $(X_n)$  une suite de v.a.i.i.d  $\mathcal{B}(p)$  alors  $M_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} p$
- Estimation d'une proportion inconnue

## Theorem

Soit  $(X_n)$  une suite de v.a.i.i.d et  $L^1$  alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_0]$$

## Theorem

**Loi Forte des Grands Nombres:** Soit  $(X_n)$  une suite de v.a.i.i.d et  $L^1$  alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s} \mathbb{E}[X_0]$$

- Une réciproque.
- Application: méthode de Monte Carlo. Soit  $f$  une fonction mesurable telle que  $f(X_0)$  soit  $L^1$

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow{p.s} \mathbb{E}[f(X_0)]$$

Rq: ici aucune régularité n'est demandé pour  $f$  ce qui n'est pas le cas dans un certain nombre de méthodes classiques d'approximation d'intégrale.

- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  et  $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2$  sont des estimateurs de la moyenne et de la variance

## Theorem

**Théorème Centrale Limite:** Soit  $(X_n)$  une suite de v.a.i.i.d et  $L^2$  de moyenne  $m$  et de variance  $\sigma^2$ . On pose

$$S_n = \sum_{i=1}^n X_i$$

alors

$$\frac{S_n - nm}{\sqrt{n\sigma^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

- Affine la LFGN: vitesse de convergence en loi.
- Intervalle de confiance

# 9) Vecteurs Gaussiens

## Definition

Un vecteur aléatoire  $X = (X_1, \dots, X_d)^t$  est appelé vecteur gaussien si toutes combinaisons linéaires de ces coordonnées est une v.a gaussienne, c'est à dire que pour tout  $a \in \mathbb{R}^d$  la v.a

$$\langle a, X \rangle = \sum_{i=1}^d a_i X_i$$

est une v.a gaussienne

- Si  $X$  est un vecteur gaussien alors pour toute matrice  $A$  le vecteur  $AX$  est encore un vecteur gaussien

## Definition

Soit  $X = (X_1, \dots, X_d)^t$  un vecteur aléatoire gaussien on note  $K$  sa matrice de covariance définie par

$$K_{i,j} = \text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j],$$

pour tout  $i, j = 1, \dots, d$ . On notera également

$$m = \mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^t$$

le vecteur de la moyenne. On notera

$$X \sim \mathcal{N}_d(m, K)$$

- La matrice  $K$  est positive
- $\mathbb{E}[\langle a, X \rangle] = \langle a, \mathbb{E}[X] \rangle$
- $\text{Var}(\langle a, X \rangle) = \text{Var}(\sum_{i=1}^d a_i X_i) = \sum_{i,j=1}^d a_i a_j \text{Cov}(X_i, X_j) = a^t K a = \langle a, K a \rangle$

- On a en particulier

$$\phi_{\langle a, X \rangle}(t) = \exp\left(i \langle a, m \rangle t - \frac{1}{2} a^t K a t^2\right)$$

- $\phi_X(x) = \mathbb{E}[e^{i\langle x, X \rangle}] = \phi_{\langle x, X \rangle}(1)$

## Proposition

*La fonction caractéristique d'un vecteur Gaussien est donnée par*

$$\phi_X(x) = \exp\left(i \langle x, m \rangle - \frac{1}{2} x^t K x\right)$$

- Les coordonnées d'un vecteur gaussien sont indépendantes si et seulement si sa matrice de covariance est diagonale



## Proposition

Soit  $X \sim \mathcal{N}_d(m, K)$  alors pour toute matrice  $A \in \mathbb{M}_{p,d}(\mathbb{R})$  alors

$$AX \sim \mathcal{N}_p(AX, AKA^t)$$

- Si  $X \sim \mathcal{N}_d(0, I_d)$  alors la loi de  $X$  est invariante par toute rotation.

- On dira qu'un vecteur gaussien  $X$  est dégénéré si sa matrice de covariance  $K$  est non inversible
- Dans le cas dégénéré, il existe donc  $a$  tel que  $Ka = 0$  ce qui implique que

$$\text{Var}(\langle a, X \rangle) = 0$$

et donc  $\langle a, X \rangle = b$  p.s donc  $X$  vit dans l'espace affine  $\{\langle a, x \rangle = b, x \in \mathbb{R}^d\}$

- Si  $K$  est inversible alors  $\sqrt{K}^{-1}(X - m) \sim \mathcal{N}(0, I_d)$
- Si  $N \sim \mathcal{N}(0, I_d)$  alors  $X = \sqrt{K}N + m \in \mathcal{N}(m, K)$

- Si  $X \sim \mathcal{N}_d(0, I_d)$  alors les coordonnées  $(X_i)_{i=1, \dots, d}$  sont i.i.d et  $X_1 \sim \mathcal{N}(0, 1)$ . Ainsi la densité de  $X$  est donnée par le produit des densités i.e

$$f_X(x_1, \dots, x_d) = \frac{1}{\sqrt{2\pi}^d} \exp\left(-\frac{1}{2} \sum_{i=1}^d x_i^2\right)$$

- Dans le cas où  $K$  est inversible on a

$$f_X(x_1, \dots, x_d) = \frac{1}{\sqrt{(2\pi)^d \det K}} \exp\left(-\frac{1}{2} \langle (x - m), K^{-1}(x - m) \rangle\right)$$

- Rq: si  $X$  est gaussien toutes ses coordonnées le sont, l'inverse n'est pas vrai en général.

## Theorem

Soit  $X^{(n)}$  une suite de vecteurs aléatoires de  $\mathbb{R}^d$  i.i.d et  $L^2$  de vecteur moyen  $m$  et de matrice de covariance  $K$ . On pose  $S^{(n)} = \sum_{i=1}^n X^{(i)}$  alors on a

$$n^{-1/2} \sqrt{K}^{-1} (S^{(n)} - nm) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, I_d)$$

ou

$$n^{-1/2} (S^{(n)} - nm) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, K)$$

Soit  $(X_n)$  i.i.d de moyenne  $m$  et de variance  $\sigma^2$

- Le TCL nous dit

$$\sqrt{n}(\bar{X}_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

- Si  $\mathbb{E}[X_1^4] < \infty$ , en posant  $\mu^4 = \mathbb{E}(X_0 - m)^4$  alors on a

$$\sqrt{n}(\bar{S}_n^2 - nm) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^4 - \sigma^4)$$

- Soit  $X = (X_1, \dots, X_d)$  un vecteur gaussien où les  $(X_i)$  sont i.i.d de loi  $\mathcal{N}(0, 1)$  alors

$$Z = \sum_{i=1}^d X_i^2$$

est une variable aléatoire appelé v.a  $\chi^2(d)$  où  $d$  est appelé degré de liberté

- Sa densité est donnée par

$$f_Z(z) = \frac{1}{2\Gamma(k/2)} z^{\frac{k}{2}-1} e^{-\frac{z}{2}} \mathbf{1}_{z \geq 0}$$

où  $\Gamma$  est la fonction gamma.

- Soit  $X \sim \mathcal{N}(0, 1)$  et  $Z \sim \chi^2(k)$  alors la v.a

$$T = \frac{X}{\sqrt{Z/k}}$$

est une variable aléatoire dite de Student de degré  $k$

- Sa densité est donnée par

$$f_T(t) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{t^2}{2}\right)^{-\frac{k+1}{2}}$$

- Soit  $X \sim \mathcal{N}(0, 1)$  et  $Z \sim \chi^2(k)$  alors la v.a

$$T = \frac{X}{\sqrt{Z/k}}$$

est une variable aléatoire dite de Student de degré  $k$

- Sa densité est donnée par

$$f_T(t) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{t^2}{2}\right)^{-\frac{k+1}{2}}$$



## Proposition

Soit  $X \sim \mathcal{N}_d(0, I_d)$  et  $\mathbb{R}^d = F_1 \oplus \dots \oplus F_k$  une décomposition en sous espaces orthogonaux avec  $\dim(F_i) = d_i$ . On note  $P_{F_i}, i = 1, \dots, k$  les projecteurs orthogonaux associés aux espaces  $F_i, i = 1 \dots, k$ . Dans ce cas les vecteurs  $P_{F_1}(X), \dots, P_{F_k}(X)$  sont des vecteurs aléatoires gaussiens indépendants. On a également

$$\|P_{F_i}(X)\| \sim \chi^2(d_i), i = 1, \dots, k$$

- C'est de l'algèbre linéaire
- On peut exprimer un résultat plus général lorsque  $X \sim \mathcal{N}(0, K)$  avec  $K$  non dégénérée en introduisant le produit scalaire relié à  $K$  i.e  $\langle a, b \rangle_K = \langle a, Kb \rangle$ .

- Considérons

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n (X_i)^2 - (\bar{X}_n)^2$$

- Par la loi forte des grands nombres cette quantité converge vers la variance de  $X_1$

- Commençons par calculer

$$\begin{aligned}\mathbb{E}[\sigma_n^2] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] - \mathbb{E}[(\bar{X}_n)^2] \\ &= \mathbb{E}(X_1^2) - \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2\right] \\ &= \mathbb{E}(X_1^2) - \frac{1}{n^2} \sum_{i,j} \mathbb{E}[X_i X_j] \\ &= \mathbb{E}(X_1^2) - \frac{1}{n^2} \left( \sum_{i=j} \mathbb{E}[(X_i)^2] + \sum_{i \neq j} \mathbb{E}[X_i] \mathbb{E}[X_j] \right) \\ &= \mathbb{E}(X_1^2) - \frac{1}{n} \mathbb{E}[X_1^2] - \frac{1}{n^2} \sum_{i \neq j} \mathbb{E}[X_1]^2 \\ &= \frac{n-1}{n} \mathbb{E}[X_1^2] - \frac{n-1}{n} \mathbb{E}[X_1]^2 = \frac{n-1}{n} \text{Var}(X_1)\end{aligned}$$

- On a donc

$$\mathbb{E}[\sigma_n^2] = \frac{n-1}{n} \text{Var}(X_1)$$

- On préfère considérer

$$S_n^2 = \frac{n}{n-1} \sigma_n^2 = \frac{1}{n-1} \sum_i^n (X_i - \bar{X}_n)^2,$$

qui est encore un estimateur de la vraie variance mais qui est non biaisé.

- Supposons que  $(X_1, \dots, X_n)^t$  soit un vecteur gaussien indépendant de loi  $\mathcal{N}(m, \sigma^2)$ . Nous avons

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

- En effet on pose  $Y = \frac{1}{\sigma}(X_1 - m, \dots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons  $F = \text{Vect}(1_n)$  avec  $1_n = (1, \dots, 1)^t$ . On voit que  $\dim(F) = 1$  et  $\dim(F^\perp) = n-1$ .
- Alors maintenant soit  $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \dots, \bar{X}_n - m)^t$  et donc

$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^t$$

- Le théorème de Cochran nous dit que  $\|P_{F^\perp}(X)\|^2 \sim \chi^2(n-1)$ . Il est facile de voir que

$$\|P_{F^\perp}(X)\|^2 = \frac{n-1}{\sigma^2} S_n^2$$

- Supposons que  $(X_1, \dots, X_n)^t$  soit un vecteur gaussien indépendant de loi  $\mathcal{N}(m, \sigma^2)$ . Nous avons

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

- En effet on pose  $Y = \frac{1}{\sigma}(X_1 - m, \dots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons  $F = \text{Vect}(1_n)$  avec  $1_n = (1, \dots, 1)^t$ . On voit que  $\dim(F) = 1$  et  $\dim(F^\perp) = n-1$ .
- Alors maintenant soit  $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \dots, \bar{X}_n - m)^t$  et donc

$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^t$$

- Le théorème de Cochran nous dit que  $\|P_{F^\perp}(X)\|^2 \sim \chi^2(n-1)$ . Il est facile de voir que

$$\|P_{F^\perp}(X)\|^2 = \frac{n-1}{\sigma^2} S_n^2$$

- Supposons que  $(X_1, \dots, X_n)^t$  soit un vecteur gaussien indépendant de loi  $\mathcal{N}(m, \sigma^2)$ . Nous avons

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

- En effet on pose  $Y = \frac{1}{\sigma}(X_1 - m, \dots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons  $F = \text{Vect}(1_n)$  avec  $1_n = (1, \dots, 1)^t$ . On voit que  $\dim(F) = 1$  et  $\dim(F^\perp) = n - 1$ .
- Alors maintenant soit  $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \dots, \bar{X}_n - m)^t$  et donc

$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^t$$

- Le théorème de Cochran nous dit que  $\|P_{F^\perp}(X)\|^2 \sim \chi^2(n-1)$ . Il est facile de voir que

$$\|P_{F^\perp}(X)\|^2 = \frac{n-1}{\sigma^2} S_n^2$$

- Supposons que  $(X_1, \dots, X_n)^t$  soit un vecteur gaussien indépendant de loi  $\mathcal{N}(m, \sigma^2)$ . Nous avons

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

- En effet on pose  $Y = \frac{1}{\sigma}(X_1 - m, \dots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons  $F = \text{Vect}(1_n)$  avec  $1_n = (1, \dots, 1)^t$ . On voit que  $\dim(F) = 1$  et  $\dim(F^\perp) = n-1$ .
- Alors maintenant soit  $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \dots, \bar{X}_n - m)^t$  et donc

$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^t$$

- Le théorème de Cochran nous dit que  $\|P_{F^\perp}(X)\|^2 \sim \chi^2(n-1)$ . Il est facile de voir que

$$\|P_{F^\perp}(X)\|^2 = \frac{n-1}{\sigma^2} S_n^2$$



- Supposons que  $(X_1, \dots, X_n)^t$  soit un vecteur gaussien indépendant de loi  $\mathcal{N}(m, \sigma^2)$ . Nous avons

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

- En effet on pose  $Y = \frac{1}{\sigma}(X_1 - m, \dots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons  $F = \text{Vect}(1_n)$  avec  $1_n = (1, \dots, 1)^t$ . On voit que  $\dim(F) = 1$  et  $\dim(F^\perp) = n - 1$ .
- Alors maintenant soit  $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \dots, \bar{X}_n - m)^t$  et donc

$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^t$$

- Le théorème de Cochran nous dit que  $\|P_{F^\perp}(X)\|^2 \sim \chi^2(n-1)$ . Il est facile de voir que

$$\|P_{F^\perp}(X)\|^2 = \frac{n-1}{\sigma^2} S_n^2$$

- Cela permet entre autre de construire des intervalles de confiance pour la variance d'une loi gaussienne. On pose  $\chi_{1-\alpha}^k$  le quantile de la loi de  $\chi^2(k)$  c'est à dire si  $T \sim \chi^2(k)$  alors

$$\mathbb{P}[\chi_{\alpha/2}^k \leq T \leq \chi_{1-\alpha/2}^k] = 1 - \alpha$$

- On a alors

$$\mathbb{P}\left[\chi_{\alpha/2}^k \leq \frac{n-1}{\sigma^2} S_n^2 \leq \chi_{1-\alpha/2}^{n-1}\right] = 1 - \alpha$$

- Cela implique

$$\mathbb{P}\left[\frac{n-1}{\chi_{1-\alpha/2}^{n-1}} S_n^2 \leq \sigma^2 \leq \frac{n-1}{\chi_{\alpha/2}^{n-1}} S_n^2\right] = 1 - \alpha$$

et l'intervalle

$$\left[\frac{n-1}{\chi_{1-\alpha/2}^{n-1}} S_n^2, \frac{n-1}{\chi_{\alpha/2}^{n-1}} S_n^2\right]$$

est un intervalle de confiance au degré  $\alpha$  pour la vraie variance  $\sigma^2$  of  $X_1$ .

- Quand  $X_1, \dots, X_n$  est Gaussien indépendant  $\mathcal{N}(m, \sigma^2)$  alors

$$\sqrt{n} \left( \frac{\bar{X}_n - m}{\sigma} \right) \sim \mathcal{N}(0, 1)$$

alors si  $\sigma^2$  est connu cela permet de construire un IC pour  $\mu$

- Si  $\sigma^2$  est inconnu on remplace  $\sigma$  par  $S_n$  et on a

$$\sqrt{n} \left( \frac{\bar{X}_n - m}{S_n} \right) \sim \mathcal{T}_{n-1}$$

où  $\mathcal{T}_{n-1}$  suit une loi de Student avec  $n - 1$  degré de liberté.

# Test du $\chi^2$

Test d'adéquation du  $\chi^2$ :

- On observe une variable aléatoire  $X$  de support connu  $S(X) = \{a_1, \dots, a_r\}$  et  $p_j = \mathbb{P}(X = a_j) = Q(\{a_j\}), j = 1, \dots, r$  inconnues. On note  $p = (p_1, \dots, p_r)$  le vecteur des probabilités
- On pose  $Q_0 = \sum_i \pi_i \delta_i$  de même support mais avec un vecteur  $\pi = (\pi_1, \dots, \pi_r)$  connu où  $\pi_i > 0$
- On cherche à tester l'hypothèse  $H_0 : Q = Q_0$  contre  $H_1 : Q \neq Q_0$ .
- On observe un vecteur  $(X_n)$  une suite de v.a.i.i.d de loi  $Q$ . Pour  $n \in \mathbb{N}$ , on pose

$$N_j = \sum_{i=1}^n \mathbf{1}_{X_i=a_j}$$

- On peut voir que le vecteur  $N = (N_1, N_2, \dots, N_r)^t$  suit une loi multinomiale  $\mathcal{M}(n, p_1, \dots, p_r)$  i.e

$$\mathbb{P}(N_1 = n_1, \dots, N_r = n_r) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}, \quad n_1 + \dots + n_r = n$$

- On pose

$$T_n = \sum_{j=1}^r \frac{(N_j - n\pi_j)^2}{n\pi_j}$$

- On voit que sous  $H_0$  cette quantité va être proche de 0 alors que sous  $H_1$  elle va exploser.

## Theorem

- Sous l'hypothèse  $H_0$  on a

$$T_n \xrightarrow{\mathcal{L}} \chi^2(r-1)$$

- Sous l'hypothèse  $H_1$  on a

$$T_n \xrightarrow{p.s} +\infty$$

- Test d'homogénéité, test d'indépendance

- Soit  $(X_n)$  une séquence de v.a i.i.d  $L^2$ . Notons  $\theta = \mathbb{E}[X_1]$  et  $\sigma^2 = \text{Var}(X_1)$ . Rappelons que si

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

alors le TCL dit

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

- Dans certaine situation on considère  $f$  et on veut regarder

$$\sqrt{n}(f(\bar{X}_n) - f(\theta))$$

- On utilise la méthode appelée Delta méthode

- Soit  $(X_n)$  une séquence de v.a i.i.d  $L^2$ . Notons  $\theta = \mathbb{E}[X_1]$  et  $\sigma^2 = \text{Var}(X_1)$ . Rappelons que si

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

alors le TCL dit

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

- Dans certaine situation on considère  $f$  et on veut regarder

$$\sqrt{n}(f(\bar{X}_n) - f(\theta))$$

- On utilise la méthode appelée Delta méthode

- Gardons à l'esprit le résultat du TCL

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

- Considérons  $f(x) = ax + b$  alors on a

$$\sqrt{n}(a\bar{X}_n - a\theta) \xrightarrow{\mathcal{L}} a\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, a^2\sigma^2)$$

- Supposons que  $f$  est différentiable en  $\theta$  on peut écrire  $f(x) = f(\theta) + f'(\theta)(x - \theta) + o(|x - \theta|)$ . Puisque  $\bar{X}_n - \theta$  converge vers 0 presque sûrement, on a convergence en proba vers 0. On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + o_{\mathbb{P}}(|\bar{X}_n - \theta|)$$



- Gardons à l'esprit le résultat du TCL

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

- Considérons  $f(x) = ax + b$  alors on a

$$\sqrt{n}(a\bar{X}_n - a\theta) \xrightarrow{\mathcal{L}} a\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, a^2\sigma^2)$$

- Supposons que  $f$  est différentiable en  $\theta$  on peut écrire  $f(x) = f(\theta) + f'(\theta)(x - \theta) + o(|x - \theta|)$ . Puisque  $\bar{X}_n - \theta$  converge vers 0 presque sûrement, on a convergence en proba vers 0. On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + o_{\mathbb{P}}(|\bar{X}_n - \theta|)$$

- Gardons à l'esprit le résultat du TCL

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

- Considérons  $f(x) = ax + b$  alors on a

$$\sqrt{n}(a\bar{X}_n - a\theta) \xrightarrow{\mathcal{L}} a\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, a^2\sigma^2)$$

- Supposons que  $f$  est différentiable en  $\theta$  on peut écrire  $f(x) = f(\theta) + f'(\theta)(x - \theta) + o(|x - \theta|)$ . Puisque  $\bar{X}_n - \theta$  converge vers 0 presque sûrement, on a convergence en proba vers 0. On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + o_{\mathbb{P}}(|\bar{X}_n - \theta|)$$

- On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + o_{\mathbb{P}}(|\bar{X}_n - \theta|)$$

into  $\sqrt{n}(f(\bar{X}_n) - f(\theta))$ , et on a

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) = \sqrt{nf'(\theta)}(\sqrt{n}(\bar{X}_n - \theta))(1 + o_{\mathbb{P}}(1))$$

- Maintenant  $1 + o_{\mathbb{P}}(1)$  converge vers 1 en proba et donc en loi. On peut alors utiliser le [lemme de Slutsky](#) et on a

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) \xrightarrow{\mathcal{L}} f'(\theta)\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, f'(\theta)^2\sigma^2)$$

- On peut alors écrire

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + o_{\mathbb{P}}(|\bar{X}_n - \theta|)$$

into  $\sqrt{n}(f(\bar{X}_n) - f(\theta))$ , et on a

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) = \sqrt{n}f'(\theta)(\sqrt{n}(\bar{X}_n - \theta))(1 + o_{\mathbb{P}}(1))$$

- Maintenant  $1 + o_{\mathbb{P}}(1)$  converge vers 1 en proba et donc en loi. On peut alors utiliser le [lemme de Slutsky](#) et on a

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) \xrightarrow{\mathcal{L}} f'(\theta)\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, f'(\theta)^2\sigma^2)$$

# **10) Estimateur du maximum de vraisemblance**

Le contexte est le suivant

- On observe un échantillon  $(X_1, \dots, X_n)$  de loi  $\mathcal{L}(\theta)$ . On connaît à priori la forme des lois  $\mathcal{L}(\theta)$  mais on ne connaît pas  $\theta$  (exemple paramètre d'une loi de Bernoulli).
- On cherche alors à définir un estimateur de ce paramètre c'est à dire une fonction

$$\hat{\theta}_n : (x_1, \dots, x_n) \rightarrow \hat{\theta}_n(x_1, \dots, x_n)$$

tel que

$$\lim_{n \rightarrow \infty} \hat{\theta}_n(X_1, \dots, X_n) = \theta \text{ p.s}$$

- Il peut exister plusieurs estimateurs satisfaisant cette contrainte, on va chercher celui qui maximise la vraisemblance

$$L_n(x_1, \dots, x_n, \theta)$$

Cas discret: soit  $(X_1, \dots, X_n)$  un échantillon de loi  $\mathbb{P}_\theta$

- Dans le cas discret la vraisemblance est décrite par

$$L_n(x_1, \dots, x_n, \theta) = \mathbb{P}_\theta(X_1 = x_1) \dots \mathbb{P}_\theta(X_n = x_n)$$

- On pose alors estimateur du maximum de vraisemblance

$$\hat{\theta}_n = \operatorname{argmax}_\theta(L_n(x_1, \dots, x_n, \theta))$$

- On est alors amené souvent à déterminer ce maximum en résolvant

$$\frac{d}{d\theta} L_n(x_1, \dots, x_n, \theta) = 0$$

- Exemple:  $\mathcal{B}(p)$ ,  $\mathcal{P}(\lambda)$

Cas continu: soit  $(X_1, \dots, X_n)$  un échantillon de densité  $f_\theta(\cdot)$

- Dans le cas discret la vraisemblance est décrite par

$$L_n(x_1, \dots, x_n, \theta) = f_\theta(x_1) \dots f_\theta(x_n)$$

- On pose alors estimateur du maximum de vraisemblance

$$\hat{\theta}_n = \operatorname{argmax}_\theta(L_n(x_1, \dots, x_n, \theta))$$

- On est alors amené souvent à déterminer ce maximum en résolvant

$$\frac{d}{d\theta} L_n(x_1, \dots, x_n, \theta) = 0$$

- Exemple:  $\mathcal{U}([0, a])$ ,  $\mathcal{E}(\lambda)$ ,  $\mathcal{N}(m, \sigma^2)$ .