Probabilité Agrégation

Clément Pellegrini

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 Ω un ensemble

Definition

$\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu (ou σ-algèbre) sur Ω 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 Ω est discret ou dénombrable
3. Lorsque Ω 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 Ω un ensemble

Definition

$\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu (ou σ-algèbre) sur Ω 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 Ω est discret ou dénombrable
3. Lorsque Ω 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

Definition

Soit X et Y deux ensembles munis de tribus $\mathcal{A}$ pour X et $\mathcal{B}$ pour Y. Une application $f : (X, \mathcal{A}) \to (Y, \mathcal{B})$ 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 Ω muni d'une tribu $\mathcal{A}$ s'appelle un espace mesurable on le note $(\Omega, \mathcal{A})$

Definition

Une mesure μ sur $(\Omega, \mathcal{A})$ est une application de $\mathcal{A} \to [0, +\infty]$ 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
$$\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

1. Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré.
2. **Lorsque** μ **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 μ 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 Ω muni d'une tribu $\mathcal{A}$ s'appelle un espace mesurable on le note $(\Omega, \mathcal{A})$

Definition

Une mesure μ sur $(\Omega, \mathcal{A})$ est une application de $\mathcal{A} \to [0, +\infty]$ 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
$$\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

1. Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré.
2. **Lorsque μ 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 μ 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

On rappelle que

Proposition

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* $A \subset B$*:* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$
4. *Si* (A_n) *est une suite croissante i.e* $A_n \subset A_{n+1}$

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_n A_n)$$

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

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_n A_n)$$

Définitions

Definition

$\mathcal{M} \subset \mathcal{P}(\Omega)$ est une *classe monotone* (c.m. en abrégé) si:

1. $\Omega \in \mathcal{M}$,
2. $\left[A, B \in \mathcal{M} \text{ et } B \subset A\right] \Rightarrow A \setminus B = A \cap B^c \in \mathcal{M}$,
3. $\left[\forall i \in \mathbb{N},\ A_i \in \mathcal{M} \text{ et } A_i \subset A_{i+1}\right] \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{M}$.

- Toute tribu est une classe monotone
- Soient μ et ν deux mesures. L'ensemble $\{A \mid \mu(A) = \nu(A)\}$ est une classe monotone

Définitions

Definition

$\mathcal{M} \subset \mathcal{P}(\Omega)$ est une *classe monotone* (c.m. en abrégé) si:

1. $\Omega \in \mathcal{M}$,
2. $\left[A, B \in \mathcal{M} \text{ et } B \subset A\right] \Rightarrow A \setminus B = A \cap B^c \in \mathcal{M}$,
3. $\left[\forall i \in \mathbb{N},\ A_i \in \mathcal{M} \text{ et } A_i \subset A_{i+1}\right] \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{M}$.

- Toute tribu est une classe monotone
- Soient μ et ν deux mesures. L'ensemble $\{A \mid \mu(A) = \nu(A)\}$ est une classe monotone

Définitions

Definition

Une intersection quelconque de c.m. est une c.m.

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$. La classe monotone engendrée par $\mathcal{E}$, notée $\mathcal{M}(\mathcal{E})$, est par définition l'intersection de toutes les c.m. contenant $\mathcal{E}$.
C'est la plus petite c.m. sur Ω qui contient $\mathcal{E}$.

Proposition

Soit $\mathcal{M}$ une classe monotone. Si elle est stable par intersection (ou union) finie, alors c'est une tribu.

Theorem (Classes monotones)

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$ stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Définitions

Definition

Une intersection quelconque de c.m. est une c.m.

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$. La classe monotone engendrée par $\mathcal{E}$, notée $\mathcal{M}(\mathcal{E})$, est par définition l'intersection de toutes les c.m. contenant $\mathcal{E}$.
C'est la plus petite c.m. sur Ω qui contient $\mathcal{E}$.

Proposition

Soit $\mathcal{M}$ une classe monotone. Si elle est stable par intersection (ou union) finie, alors c'est une tribu.

Theorem (Classes monotones)

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$ stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Définitions

Definition

Une intersection quelconque de c.m. est une c.m.

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$. La classe monotone engendrée par $\mathcal{E}$, notée $\mathcal{M}(\mathcal{E})$, est par définition l'intersection de toutes les c.m. contenant $\mathcal{E}$.
C'est la plus petite c.m. sur Ω qui contient $\mathcal{E}$.

Proposition

Soit $\mathcal{M}$ une classe monotone. Si elle est stable par intersection (ou union) finie, alors c'est une tribu.

Theorem (Classes monotones)

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$ stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Définitions

Definition

Un ensemble $\mathcal{H}$ de fonctions de $\Omega \to \mathbb{R}$ est dit stable par convergence monotone bornée si pour toute suite $(f_n)_{n\in\mathbb{N}}$, croissante et bornée $(\exists\ C,\ \forall n,\ \forall \omega,\ \left|f_n(\omega)\right| \leqslant C)$ de fonctions de $\mathcal{H}$, la limite $f = \lim f_n$ est dans $\mathcal{H}$.

Theorem

(Version fonctionnelle du théorème de classe monotone)
Soit $\mathcal{H}$ un $\mathbb{R}$-espace vectoriel de fonctions bornées de $\Omega \to \mathbb{R}$ stable par convergence monotone bornée et contenant les constantes. Soit $\mathcal{E} \subset \mathcal{H}$ un ensemble de fonctions stable par multiplication.
Alors $\mathcal{H}$ contient l'ensemble $b\left(\sigma(\mathcal{E})\right)$ des fonctions bornées mesurables pour

$$\sigma(\mathcal{E}) = \sigma\left(\left\{f^{-1}(B),\ f \in \mathcal{E},\ B \in \mathcal{B}(\mathbb{R})\right\}\right).$$

Définitions

Definition

Un ensemble $\mathcal{H}$ de fonctions de $\Omega \to \mathbb{R}$ est dit stable par convergence monotone bornée si pour toute suite $(f_n)_{n\in\mathbb{N}}$, croissante et bornée $(\exists\ C,\ \forall n,\ \forall\omega,\ \left|f_n(\omega)\right| \leqslant C)$ de fonctions de $\mathcal{H}$, la limite $f = \lim f_n$ est dans $\mathcal{H}$.

Theorem

(Version fonctionnelle du théorème de classe monotone)
Soit $\mathcal{H}$ un $\mathbb{R}$-espace vectoriel de fonctions bornées de $\Omega \to \mathbb{R}$ stable par convergence monotone bornée et contenant les constantes. Soit $\mathcal{E} \subset \mathcal{H}$ un ensemble de fonctions stable par multiplication.
Alors $\mathcal{H}$ contient l'ensemble $b\left(\sigma(\mathcal{E})\right)$ des fonctions bornées mesurables pour

$$\sigma(\mathcal{E}) = \sigma\left(\left\{f^{-1}(B),\ f \in \mathcal{E},\ B \in \mathcal{B}(\mathbb{R})\right\}\right).$$

Définitions

Theorem

(Caractérisation des mesures finies)
Soit $(\Omega, \mathcal{A})$ *un espace mesuré. Soit* $\mathcal{E} \subset \mathcal{A} \subset \mathcal{P}(\Omega)$ *telle que*

- $\Omega \in \mathcal{E}$,
- $A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}$,
- $\sigma(\mathcal{E}) = \mathcal{A}$.

Si μ et ν *sont deux mesures finies sur* $(\Omega, \mathcal{A})$ *qui coincident sur* $\mathcal{E}$ *alors* $\mu \;=\; \nu$.

Définitions

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 Ω 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 $X : \Omega \to \mathbb{R}$ 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 λ.

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 Ω 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 $X : \Omega \to \mathbb{R}$ 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 λ.

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 Ω qui rend l'application $X : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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 $X_1, \ldots, X_n$ et on note $\sigma(X_1, \ldots, X_n)$ 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 Ω qui rend l'application $X : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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 $X_1, \ldots, X_n$ et on note $\sigma(X_1, \ldots, X_n)$ la plus petite tribu rendant les X_i mesurables.

Proposition (Doob)

Soit $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$ *et* $Y : \Omega \to \mathbb{R}^d$ *deux v.a.. Alors* Y *est* $\sigma(X)$*-mesurable si et seulement s'il existe une fonction* $h : E \to \mathbb{R}^d$ *borélienne avec* $Y = h(X)$.

Définitions

Definition

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

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

Proposition

1. *F_X est croissante*
2. $\lim_{t \to -\infty} F_X(t) = 0$ *et* $\lim_{t \to +\infty} F_X(t) = 1$
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*

Définitions

Definition

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

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

Proposition

1. *F_X est croissante*
2. $\lim\limits_{t\to-\infty} F_X(t) = 0$ *et* $\lim\limits_{t\to+\infty} F_X(t) = 1$
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*

Définitions

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

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

  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

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

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

Variable aléatoire discrète

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 $S(X) = \{a_1, \ldots, a_n\}$ 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 $\{a_1, \ldots, a_n\}$ si $p_i = \frac{1}{n}$ pour tout $i \in \{1, \ldots, n\}$ 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 $S(X) = \{a_1, \ldots, a_n\}$ 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 $\{a_1, \ldots, a_n\}$ si $p_i = \frac{1}{n}$ pour tout $i \in \{1, \ldots, n\}$ 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 succés (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 succés. 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 $S(X) = \{0, \ldots, n\}$ 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,: $X \sim \mathcal{B}(n, p)$.

Exemples usuels

- Si X modélise un succés (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 succés. 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 $S(X) = \{0, \ldots, n\}$ 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,: $X \sim \mathcal{B}(n, p)$.

Exemples usuels

- 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 géometrique** 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

Exemples usuels

- 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 géometrique** 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

Exemples usuels

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

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

V.a continues ou à densité

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 $f \geqslant 0$
- $\int_{\mathbb{R}} f(x)dx = 1$
- Pour un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ on a pour tout borélien A de $\mathbb{R}^d$:

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

V.a continues ou à densité

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 $\underline{f}$
- On a nécessairement que $f \geqslant 0$
- $\int_{\mathbb{R}} f(x)dx = 1$
- Pour un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ on a pour tout borélien A de $\mathbb{R}^d$:

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

Exemples usuels

- 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 σ^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 $a > 0$ $X \sim \mathcal{C}(a)$ 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

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

Exemples usuels

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 ϕ continue et bornée

Theorem

Soit X une v.a, pour toute fonction continue bornée ϕ

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

- Dans le cas discret $\mathbb{E}[\phi(X)] = \sum_{x \in S(X)} \phi(x)\mathbb{P}[X = x]$
- 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 ϕ 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.

Formule du transfert

Definition

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

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

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

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

Formule du transfert

- Inégalité de Markov

$$\mathbb{P}(|X| \geqslant t) \leqslant \frac{\mathbb{E}(|X|)}{t}$$

- Inégalité de Bienaymé-Tchebychev

$$\mathbb{P}(|X - \mathbb{E}(X)| \geqslant t) \leqslant \frac{Var(X)}{t^2}$$

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

Exemples

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

4) Transformée exponentielle

Fonction caractéristique

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

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

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

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

Fonction caractéristique

Theorem

$\phi_X = \phi_Y$ *ssi* $P_X = P_Y$

Theorem

Soit X *v.a. à valeurs dans* $\mathbb{R}^d$. *On note* $\lambda_{\mathbb{R}^d}$ *la mesure de Lebesgue sur* $\mathbb{R}^d$. *Si* $\Phi_X \in L^1(\mathbb{R}^d, \lambda_{\mathbb{R}^d})$ *alors* $P_X \ll \lambda_{\mathbb{R}^d}$ *et sa densité est donnée par*

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \xi, x\rangle} \Phi_X(\xi) d\xi.$$

Fonction caractéristique

Theorem

$\phi_X = \phi_Y$ *ssi* $P_X = P_Y$

Theorem

Soit X *v.a. à valeurs dans* $\mathbb{R}^d$. *On note* $\lambda_{\mathbb{R}^d}$ *la mesure de Lebesgue sur* $\mathbb{R}^d$. *Si* $\Phi_X \in L^1(\mathbb{R}^d, \lambda_{\mathbb{R}^d})$ *alors* $P_X \ll \lambda_{\mathbb{R}^d}$ *et sa densité est donnée par*

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \xi, x\rangle} \Phi_X(\xi) d\xi.$$

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 $\phi_X(t) = \frac{\lambda}{\lambda - it}$
- $X \sim C(a)$ alors $\phi_X(t) = exp(-a|t|)$
- $X \sim \mathcal{N}(m, \sigma^2)$ alors $\phi_X(t) = exp(imt - \frac{\sigma^2 t^2}{2})$

Fonction caractéristique et moments

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 ϕ_X est $p \geqslant 2$ fois dérivable en 0 alors X admet des moments d'ordre $2[p/2]$

Autres transformées

- 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^∞ 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 $G_X'(1) = \mathbb{E}(X)$
- Par unicité du DSE en 0 la fonction génératrice des moments caractérise la loi.

Autres transformées

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

$$L_X(t) = \mathbb{E}[e^{\langle t, X \rangle}]$$

Theorem

Soient X, Y deux v.a. à valeurs dans $\mathbb{R}^d$. Soit $\epsilon > 0$. Si pour tout s tel que $||s|| \leqslant \epsilon$ on a $L_X(s) = L_Y(s) < +\infty$, alors $P_X = P_Y$.

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 $\sigma(X) \perp\!\!\!\perp \sigma(Y)$

Définition

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

Somme de v.a indépendantes

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

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

- 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.

Somme de v.a indépendantes

Definition

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

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

pour tout $A_1, \ldots A_k$ dans la tribu.
En particulier pour le cas d'une famille finie $X_1, \ldots, X_N$, cette famille est dite indépendante si:

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

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

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

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

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 $A \perp\!\!\!\perp B$ 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 joeur 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 $\mathbb{P}(A)$

Formule des probas totales et formule de Bayes

Proposition

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

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

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

Loi conditionnelles

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

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

- 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 $\mathbb{P}(Y \in \cdot|X)$

Loi conditionnelles

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

Définition

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 $\quad \mathbb{P}(\lim X_n = X) = 1$

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

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

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

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

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

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

  $\lim\limits_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ on note $X_n \xrightarrow{Loi} X$

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* φ *bornée et uniformément continue sur* E*,*

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

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

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

Convergence en loi

Pour une v.a on note F_X sa fonction de répartition et ϕ_X sa fonction caractéristique

Theorem

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

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

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*

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

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

Convergence monotone, convergence dominée, Fatou

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*

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

- *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*

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

Lien entre les convergences

On a les liens suivants

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

Borel cantelli

Montrer des convergences presque sûres? Le Lemme de Borel Canteli

Lemma

Soit $(A_n)_{n\in\mathbb{N}}$ une suite dévènements

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

$$\mathbb{P}(\limsup_{n\to\infty} A_n) := \mathbb{P}(\bigcap_{n\geqslant 0}\bigcup_{k\geqslant n} A_k) = 0$$

  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*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) := \mathbb{P}(\bigcap_{n\geqslant 0}\bigcup_{k\geqslant n} A_k) = 1$$

  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

Exemples

Des exemples et des contre-exemples

- Pour tout n on considère la variable aléatoire X_n qui prend les valeurs $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$ 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 contrexemples...)
- 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 $\mathbb{P}[X_n = a_n] = \frac{1}{n}$

Slutsky

- (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 constance 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

Objectif

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

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

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

Loi faible des grands nombres L^2 et L^1

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

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

Loi forte des grands nombres

Theorem

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

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

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

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

  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

Théorème centrale limite

Theorem

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

Définition

Definition

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

$$< a, X >= \sum_{i=1}^{d} a_i X_i$$

est une v.a gaussienne

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

Matrice de covariance

Definition

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

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

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

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

le vecteur de la moyenne. On notera

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

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

Fonction caractéristique

- On a en particulier

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

- $\phi_X(\xi) = \mathbb{E}[e^{i<\xi,X>}] = \phi_{<\xi,X>}(1)$

Proposition

La fonction caractéristique d'un vecteur Gaussien est donnée par

$$\phi_X(\xi) = \exp\left(i < \xi, m > - \frac{1}{2} \xi^t K \xi\right)$$

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

Transformation linéaire

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.

Centrer et réduire un vecteur gaussien

- 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

$$Var(< a, X >) = 0$$

  et donc $< a, X >= b$ p.s donc X vit dans l'espace affine $\{< a, x >= b, x \in \mathbb{R}^d\}$
- 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 $X = \sqrt{K}N + m \in \mathcal{N}(m, K)$

Densité

- 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

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

- Dans le cas où K est inversible on a

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

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

TCL multidé

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

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

Intervalle de confiance

Soit (X_n) i.i.d de moyenne m et de variance σ^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
$$\sqrt{n}(\bar{S}_n^2 - nm) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^4 - \sigma^4)$$

Transformation

- Soit $X = (X_1, \ldots, X_d)$ 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
$$f_Z(z) = \frac{1}{2\Gamma(k/2)} z^{\frac{k}{2}-1} e^{-\frac{z}{2}} \mathbf{1}_{z \geqslant 0}$$
où Γ est la fonction gamma.

Transformation

- 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 alétoire dite de Student de degré k

- Sa densité est donnée par

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

Transformation

- 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 alétoire dite de Student de degré k

- Sa densité est donnée par

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

Théorème de Cochran

Proposition

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

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

- 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 $< a, b >_K = < a, Kb >$.

Application

- 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

Confidence set

- Commençons par calculer

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

- On a donc

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

- 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 variance mais qui est non biaisé.

Confidence set

- Supposons que $(X_1, \ldots, X_n)^t$ 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 $Y = \frac{1}{\sigma}(X_1 - m, \ldots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons $F = Vect(1_n)$ avec $1_n = (1, \ldots, 1)^t$. On voit que $dim(F) = 1$ et $dim(F^\perp) = n - 1$.
- Alors maintenant soit $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \ldots \bar{X}_n - m)^t$ et donc
$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)^t$$
- Le théorème de Cochran nous dit que $||P_{F^\perp}(X)||^2 \sim \chi^2(n-1)$. Il est facile de voir que
$$||P_{F^\perp}(X)||^2 = \frac{n-1}{\sigma^2} S_n^2$$

Confidence set

- Supposons que $(X_1, \ldots, X_n)^t$ 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 $Y = \frac{1}{\sigma}(X_1 - m, \ldots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons $F = Vect(1_n)$ avec $1_n = (1, \ldots, 1)^t$. On voit que $dim(F) = 1$ et $dim(F^\perp) = n - 1$.
- Alors maintenant soit $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \ldots \bar{X}_n - m)^t$ et donc
$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)^t$$
- Le théorème de Cochran nous dit que $||P_{F^\perp}(X)||^2 \sim \chi^2(n-1)$. Il est facile de voir que
$$||P_{F^\perp}(X)||^2 = \frac{n-1}{\sigma^2} S_n^2$$

Confidence set

- Supposons que $(X_1, \ldots, X_n)^t$ 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 $Y = \frac{1}{\sigma}(X_1 - m, \ldots, X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons $F = Vect(1_n)$ avec $1_n = (1, \ldots, 1)^t$. On voit que $dim(F) = 1$ et $dim(F^\perp) = n - 1$.
- Alors maintenant soit $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X \right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m, \ldots \bar{X}_n - m)^t$ et donc
$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)^t$$
- Le théorème de Cochran nous dit que $||P_{F^\perp}(X)||^2 \sim \chi^2(n-1)$. Il est facile de voir que
$$||P_{F^\perp}(X)||^2 = \frac{n-1}{\sigma^2} S_n^2$$

Confidence set

- Supposons que $(X_1,\ldots,X_n)^t$ 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 $Y = \frac{1}{\sigma}(X_1 - m,\ldots,X_n - m)^t \sim \mathcal{N}_n(0, I_n)$
- Posons $F = Vect(1_n)$ avec $1_n = (1,\ldots,1)^t$. On voit que $dim(F) = 1$ et $dim(F^\perp) = n-1$.
- Alors maintenant soit $P_F(X) = \left\langle \frac{1_n}{\sqrt{n}}, X\right\rangle \frac{1_n}{\sqrt{n}} = \frac{1}{\sigma}(\bar{X}_n - m,\ldots \bar{X}_n - m)^t$ et donc
$$P_{F^\perp}(X) = X - P_F(X) = \frac{1}{\sigma}(X_1 - \bar{X}_n,\ldots,X_n - \bar{X}_n)^t$$
- Le théorème de Cochran nous dit que $||P_{F^\perp}(X)||^2 \sim \chi^2(n-1)$. Il est facile de voir que
$$||P_{F^\perp}(X)||^2 = \frac{n-1}{\sigma^2}S_n^2$$

Confidence set

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

Confidence set

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

$$\mathbb{P}[\chi^k_{\alpha/2} \leqslant T \leqslant \chi^k_{1-\alpha/2}] = 1 - \alpha$$

- On a alors

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

- Cela implique

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

  et l'intervalle

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

  est un intervalle de confiance au degré α pour la variance σ^2 of X_1.

Confidence set

- Quand $X_1, \ldots, X_n$ 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 σ^2 est connu cela permet de construire un IC pour μ
- Si σ^2 est inconnu on remplace σ 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 χ^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 mutinomiale $\mathcal{M}(n, p_1, \dots, p_r)$ i.e

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

Test du *chi*^{2}

- 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

Delta method

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

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

- On utilise la méthode appelée Delta méthode

Delta method

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

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

- On utilise la méthode appelée Delta méthode

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

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

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

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

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

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

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

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

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

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

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

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

Delta method

- On peut alors écrire

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

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

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

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

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

Delta method

- On peut alors écrire

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

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

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

- Maintenant $1 + \circ_{\mathbb{P}}(1)$ 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

Heuristique

Le contexte est le suivant

- On observe un échantillon $(X_1, \ldots, X_n)$ de loi $\mathcal{L}(\theta)$. On connaît à priori la forme des lois $\mathcal{L}(\theta)$ mais on ne connaît pas θ (exemple paramètre d'une loi de Bernoulli).
- On cherche alors à définir un estimateur de ce paramètre c'est à dire une fonction
$$\hat{\theta}_n : (x_1, \ldots, x_n) \to \hat{\theta}_n(x_1, \ldots, x_n)$$
tel que
$$\lim_{n\to\infty} \hat{\theta}_n(X_1, \ldots, X_n) = \theta \; p.s$$
- Il peut exister plusieurs estimateurs satisfaisant cette contrainte, on va chercher celui qui maximise la vraisemblance
$$L_n(x_1, \ldots, x_n, \theta)$$

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

$$\hat{\theta}_n = argmax_\theta(L_n(x_1, \dots, x_n, \theta))$$

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

V.a continues

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

- Dans le cas discret la vraisemblance est décrite par

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

- On pose alors estimateur du maximum de vraisemblance

$$\hat{\theta}_n = argmax_\theta(L_n(x_1, \ldots, x_n, \theta))$$

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

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

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